

중장기 대형 SW R&D 과제후보 발굴 및 추진체계

Mid/long-term SW R&D project candidates discovery
and promotion system

임영모 책임연구원

최민석 선임연구원(한국전자통신연구원)

2018.04.

공동연구기관 : 한국전자통신연구원

이 보고서는 2017년도 과학기술정보통신부 정보통신진흥기금을
지원받아 수행한 연구결과로 보고서 내용은 연구자의 견해이며,
과학기술정보통신부의 공식입장과 다를 수 있습니다.

목 차

제1장 서론	1
제1절 연구 배경	1
제2절 연구 목적 및 범위	3
제2장 쿼텀점프를 위한 R&D 추진체계	5
제1절 미국 DARPA 사례분석	5
1. 사업 기획	5
2. 조직 및 사업 운영의 특징	7
제2절 R&D 추진체계 개선방향	9
1. 사업 기획	9
2. 사업 운영 방식	11
제3장 소프트웨어 R&D 트렌드	13
제1절 기술 선도국	13
1. 미국	13
2. EU	17
제2절 연구자 커뮤니티	20
1. IEEE	20
2. ACM	23
3. 기타 저널의 특집호(Special Issues) 주제	24
제4장 중장기 대형 R&D과제 발굴 방법론	26
제1절 기존 유망기술 선정 방법론	26
제2절 본 연구의 과제 발굴 방법	28
1. 기본 방향	28
2. 과제 발굴 프로세스	29
3. 추진 체계	33

제5장 중장기 대형 SW R&D 과제 발굴	34
제1절 후보 Pool 구성	34
1. DARPA	34
2. IARPA	36
3. NSF	37
4. EU	39
제2절 예비 후보 선발	40
1. DARPA 과제 기준 예비 후보	41
2. IARPA 과제 기준 예비 후보	42
3. NSF 과제 기준 예비 후보	42
4. H2020 과제 기준 예비 후보	51
제3절 후보 선정	58
1. 예비 후보 중 후보 선정	58
2. 절대 비교 대상 후보	60
제4절 최종 후보 선정	63
1. 1차 후보 검토	64
2. 2차 후보 검토	65
3. 3차 후보 검토	66
제6장 최종 후보별 기획방향	68
제1절 멀티소스·멀티모달 기계학습 및 다중 가설 자동 생성	68
1. DARPA AIDA 사업	68
2. 국내 관련 연구과제	72
3. 향후 기획 방향	74
제2절 다중카메라 환경에서의 실시간 행위 인식 및 탐지	76
1. IARPA의 DIVA와 DARPA의 VIRAT	76
2. 국내 관련 연구	80
3. 향후 기획 방향	81

제3절 계산 커널의 병렬 자동 튜닝 오픈소스 소프트웨어	84
1. NSF BONSAI	84
2. 국내 관련 연구	85
3. 향후 기획 방향	87
제4절 오픈소스 소프트웨어 커뮤니티 활동 및 코드 자동 학습	90
1. H2020 CROSSMINER와 DARPA MUSE	90
2. 국내 관련 연구	94
3. 향후 기획 방향	95
제7장 결론	97
참고문헌	1

표 목 차

<표 1> 우리나라 소프트웨어 분야 기술수준	1
<표 2> DARPA의 사업 선정 기준	6
<표 3> IEEE Future Direction Initiatives	21
<표 4> 2018년도 Proceedings of the IEEE 특집호 주제	22
<표 5> IEEE Software 특집호 주제	23
<표 6> ACM Journal on Emerging Technologies in Computing 특집호 주제 ...	24
<표 7> Future Generation Computer 특집호 주제	24
<표 8> Future Generation Computer 특집호 주제	25
<표 9> 중장기 대형 SW과제 발굴 프로세스	29
<표 10> 본 연구의 약식 쌍대 비교의 응답지 양식과 응답결과 예시	32
<표 11> 대표적인 DARPA의 SW관련 사업 목록	35
<표 12> 현재 진행 중인 IAPRA 사업	36
<표 13> NSF의 수주기관 및 지원방식별 지원 금액 및 비중	37
<표 14> NSF CISE부서의 주요 사업	38
<표 15> 예비 후보 선발 대상 및 선발 비율	40
<표 16> DARPA 예비 후보의 수행 기간과 예산	41
<표 17> IARPA 예비 후보의 사업 공고와 수행 기간	42
<표 18> NSF CPS 예비 후보의 수행 기간과 예산, 주관기관	43
<표 19> NSF 빅데이터 예비 후보의 수행 기간과 예산, 주관기관	45
<표 20> NSF HPC 예비 후보의 수행 기간과 예산, 주관기관	47
<표 21> NSF 클라우드 컴퓨팅 예비 후보의 수행 기간과 예산, 주관기관	48
<표 22> NSF OSS 예비 후보의 수행 기간과 예산, 주관기관	50
<표 23> H2020 CPS 예비 후보의 수행 기간과 예산, 주관기관	52

<표 24> H2020 빅데이터 예비 후보의 수행 기간과 예산, 주관기관	53
<표 25> H2020 HPC 예비 후보의 수행 기간과 예산, 주관기관	54
<표 26> H2020 클라우드 컴퓨팅 예비 후보의 수행 기간과 예산, 주관기관	56
<표 27> H2020 OSS 예비 후보의 수행 기간과 예산, 주관기관	57
<표 28> 후보선정 기준	58
<표 29> 중장기 대형 R&D과제 발굴 과정을 통해 선정한 후보	59
<표 30> 현재 국내에서 수행중인 중장기 대형 소프트웨어 과제	61
<표 31> 최종 후보 선정 기준	63
<표 32> 1차 후보 검토 결과	64
<표 33> 2차 후보 검토에서의 전문가 의견 종합	66
<표 34> 3차 후보 검토에서의 전문가 의견 종합	67
<표 35> 데이터 세부과제를 제외한 4개 세부 과제의 입출력 관계	71
<표 36> 객체와 행위 탐지 측정 지표 기준 단계별 목표	79

그 립 목 차

<그림 1> 과학기술 발전에서의 DARPA 역할	5
<그림 2> DARPA 신규사업 기획 프로세스	7
<그림 3> 자동화된 프로그래밍 개념도	15
<그림 4> KISTEP의 유망기술 선정과정	26
<그림 5> AHP의 적용 사례: 3명의 후보 중 리더 선택	31
<그림 6> EU H2020 내 ICT 관련 사업	39
<그림 7> DARPA AIDA사업 전과 후의 정보분석의 차이	69
<그림 8> DARPA AIDA사업의 구성	70
<그림 9> AIDA와 LORELEI, DEFT의 관계	71
<그림 10> DARPA VIRAT사업의 개념과 운영사례	76
<그림 11> IARPA DIVA사업의 연구 목표	78
<그림 12> DARPA DIVA사업의 진행 단계별 연구 내용	78
<그림 13> DIVA의 평가 일정 및 6개월 단위 스프린트 수행체계	83
<그림 14> NSF BONSAI의 구성과 프로세스	84
<그림 15> Ktune의 구성	86
<그림 16> CROSSMINER 접근법	90
<그림 17> CROSSMINER 플랫폼 초기 화면 및 세부 화면	91
<그림 18> MUSE 시스템 개념도	92
<그림 19> DARPA MUSE 사업의 세부 과제 구성	93

요 약 문

본 연구는 선진국과의 소프트웨어 기술 격차를 획기적으로 줄이기 위한 퀀텀점프 전략의 추진 방안을 찾아보는데 있다. 퀀텀점프 전략을 추진하기 위해서는 과제 발굴에서 **사업 수**행까지 R&D 프로세스 전체 과정에서 기존과는 다른 방법을 사용해야 한다. 이를 위해 본 연구에서는 R&D 프로세스 전반에 걸친 개선점 도출과 중장기 대형 SW R&D 과제 후보 발굴을 위한 방법론을 개발하는데 중점을 뒀다.

새로운 R&D 추진체계를 모색하기 위해 미국 DARPA 프로그램을 살펴보고 현재 정부 R&D 기획 및 수행체계와 비교 분석했다. 퀀텀점프를 위한 R&D 추진체계 개선 방향은 다음과 같다.

① **(실질 수요에 기반한 과제 기획)** DARPA는 구체적인 군대의 실질 수요와 과학기술의 혁신적 진보를 연결하는데 초점을 두고 사업을 기획한다. 반면 국내 과제는 연구 목표를 실제 사용하는 것에 두고 있지 않아서 세부 과제간 유기적 연계가 약하고 역량 결집이 어려운 문제점을 가지고 있다.

② **(하향식(Top-Down) 과제 기획을 강화)** DARPA에서는 PM이 주도적인 역할을 하고 연구자들은 기술 전문가로서 해결 대안 탐색에 집중한다. 우리나라에서는 연구자들이 해결 대결의 탐색뿐만 아니라 기술의 사용처와 그 수요까지 함께 제안한다.

③ **(상시적 과제 기획 운영)** 현재 국내 소프트웨어를 비롯한 정보통신 분야의 과제 기획은 과제기획위원회를 중심으로 이루어진다. 과제기획위원회는 통상 특정 시점, 주로 연말에 1~2달 정도의 제한된 기간 내에 운영된다. 짧은 기간에도 불구하고 후보 발굴과 확정, 비교 검토, 선정, 상세 기획까지 빠르게 진행된다. 그러다 보니 기술 격차 극복을 위해, 특히 기술 선도국의 움직임을 포괄적으로 검토하기 어렵다.

④ **(CP와 과제책임자의 역할 재정립)** 우리나라에서 소프트웨어 관련 R&D 과제기획은 정보통신진흥센터(IITP)에 소속된 CP(Creative Planner)를 중심으로 이루어진다. CP가 과제기획을 하지만 과제책임자 선발 및 운영에는 관여하고 있지 않아 DARPA PM의 권한 및 재량과는 크게 차이가 난다. 그래서 CP는 DARPA의 Director 역할을 담당하여 과제를 기획하고 추진할 최고 수준의 인력을 선발하고 지원하는 업무를 담당한다. PM 역할을 담당할 총괄 과

제책임자는 DARPA처럼 직접 연구를 수행하지 않고, CP실에 소속되어 과제 기획과 운영에만 전념하도록 하는 방안이 필요하다.

⑤ **(과제 운영의 유연성 확보)** 기획단계에서 불확실성을 최소화할 수 있도록 1년 정도의 탐색 연구 시간 및 예산을 지원해야 한다. 또한, 과제 성격에 맞게 과제책임자가 단독 수의 계약, 공모, 경쟁 방식 등 모든 방식을 용인하여 현장 중심(Street-level)의 연구 수행이 될 수 있도록 노력하여야 한다. 과제책임자는 세부 과제 간의 전략적 보완관계를 끊임없이 모니터링하고 언제나 과제의 통합 및 종료 가능성이 가능하도록 하는 것이 필요하다.

미국과 유럽연합에서 현재 수행중인 과제를 대상으로 아래 표와 같이 4단계로 구성된 프로세스를 통해 향후 우리나라에서 획기적인 기술격차 축소를 위해 추진할만하다고 판단되는 4개의 과제 후보를 발굴했다. 이때 검토 범위를 기술 선도국의 과제로 한정된 것은 본 연구의 목적이 기술격차 축소에 있기 때문에 구체적인 과제 단위에서 목표점을 찾기 위함이다. 또한 이들 과제는 이미 해당 국가에서 사업성, 파급효과 등을 검증한 것으로 평가할 때 기술적인 부분만 집중할 수 있는 장점이 있다.

중장기 대형 SW과제 발굴 프로세스

구분		그룹	담당
1	후보 Pool 구성	미국: DARPP, IARPA, NSF 유럽연합: H2020	연구팀
2	예비 후보 선발	기관 전체: DARPA, IARPA 기술 분야 구분: NSF, H2020	연구팀
3	후보 선정	기관 전체: DARPA, IARPA 기술 분야 구분: NSF, H2020	전문가 평가 (분야당 3~4인)
4	최종 후보 선정	1차: DARPA + IARPA 2차: 1차 + NSF (단, OSS는 3차에 포함) 3차: 2차 + H2020 + 기존 과제	후보 검토 위원회 (10명)

우선 미국의 DARPA, IARPA, NSF와 EU의 Horizon2020 프로그램에서 현재 수행하고 있는 과제 중 SW와 관련 있는 3,322개를 뽑아 후보 Pool을 구성했다. 후보 Pool을 대상으로 DARPA, IARPA는 기관별, 과제 수가 많은 NSF, Horizon2020은 각각 CPS(Cyber-Physical

Systems), 빅데이터 분석, 클라우드 컴퓨팅, HPC(High-Performance Computing), 오픈 소스 소프트웨어(OSS)로 나누어 총 12개 그룹별로 예비 후보선발을 위한 검토를 진행했다. 각 그룹별로 적게는 4개 많게는 14개까지 선발하여 총 108개 과제가 예비 후보로 압축됐다.

예비 후보는 다시 12개 분야별로 3~4인으로 전문가 그룹을 구성하여 중장기 대형 소프트웨어 R&D과제 기획 대상으로 적합한 여부의 의견을 수렴했다. 적극 추천과 배재 희망의 양 극단에 해당하는지만 표현하게 했다. 기본적으로 최소 2명 이상의 전문가가 적극적으로 추천하고 동시에 한 명이라도 배재를 희망하지 않은 것들을 후보로 선정하여, 총 21개의 과제가 최종 평가를 위한 후보가 되었다.

최종 후보 선정은 10명의 후보 검토위원회를 구성하고 21개 후보들을 3차에 걸친 비교 검토를 통해 진행했다. 최종 후보 선정은 후보들 간의 쌍대 비교를 통해 이루어졌다. 쌍대 비교란 비교 대상들을 짝을 지어 둘 중 하나의 선호를 결정하는 비교 방식이다. 그런데 AHP를 비롯한 쌍대비교에서는 후보의 수가 하나씩 증가할수록 쌍대비교의 개수가 기하급수적으로 증가한다. 그래서 3단계로 나누어 진행했으며 아래 표와 같이 본 연구에서 개발한 약식 쌍대비교 방식을 사용했다.

본 연구의 약식 쌍대 비교의 응답지 양식과 응답결과 예시

더 높다	3.0 배	2.5 배	2.0 배	1.8 배	1.6 배	1.4 배	1.2 배	1.1 배	동 일	1.1 배	1.2 배	1.4 배	1.6 배	1.8 배	2.0 배	2.5 배	3.0 배	더 높다
후보1											○							후보2
후보2															○			후보3
후보3					○													후보4
후보4							○											후보5
후보5							○											후보1

1차에서는 DARPA와 IARPA의 후보를 묶어 4개 과제를 선발했으며, 2차에서는 1차에서 선발된 4개 과제와 NSF 후보들을 비교하여 4개 과제를 선발했다. 3차에서는 2차에서 선발한 4개 과제와 Horizon2020 후보들을 비교했다. 이때 최종적으로 선정된 후보가 경쟁 후보들보다 상대적으로 우수한 것뿐만 아니라 절대적으로 우리나라에서 중장기 대형 과제로 추진할 가치가 있어야 한다. 그래서 절대 비교 대상으로 현재 국내에서 진행되고 있는 중장기 대형 과제 중 하나를 선정하여 함께 비교했다. 12개 후보 중 국내 절대 비교 과제보다 높은 점수를 받은 4개 과제를 최종 후보로 선정했다.

최종 선정된 중장기 대형 과제 후보는 멀티소스 멀티모달 기계 학습 및 다중 가설 자동 생성(이하 3M, DARPA AIDA)과 다중카메라 환경에서의 실시간 행위 인식 및 탐지(이하 리

얼타임 딥뷰즈, IARPA DIVA), 계산 커널의 병렬 자동 튜닝 오픈소스 소프트웨어(이하 계산 커널 오토튜너, NSF BONSAD), 오픈소스 소프트웨어 활동 및 코드 자동 학습(이하 빅코드, H2020 CROSSMINER)이다.

선정된 4개의 최종 후보는 크게 두 가지로 구분할 수 있다. 첫 번째 그룹은 우리나라에서 현재 관련된 중장기 대형 과제가 진행 중에 있으나 기술 선도국이 새로운 도전적 목표를 설정한 경우이다. 3M과 리얼타임 딥뷰즈가 여기에 해당된다. 3M은 엑소브레인과 지능정보 플래그십 프로젝트, 인공지능 국가전략 프로젝트 등과 연관되어 있다. 이들 기존의 중장기 대형 과제들이 멀티소스 또는 멀티모달 중 하나만을 목표로 설정하고 있는데 반해 기술 선도국의 연구과제는 멀티소스와 멀티모달, 그리고 멀티가설까지 확대하고 결합하고 있다. 그러므로 국내 과제들을 포괄할 수 있는 과제를 새롭게 기획하여 연구자들의 협력을 강화할 필요가 있다. 리얼타임 딥뷰즈도 중장기 대형과제인 딥뷰에서 추진하는 것보다 도전적으로 연구 목표를 설정하고 있다. 기존의 과제들이 주로 단일 카메라 환경을 가정하고 있는데 반해 다중카메라 환경에서 실시간으로 위험 행위를 탐지하는 것을 목표로 삼고 있다. 따라서 기존의 딥뷰 과제와 새롭게 시작한 도전적인 개인연구들까지 아울러서 공공시설보호에 특화된 연구과제로 기획하는 것을 고려해봐야 한다.

두 번째 그룹은 전문가들이 기술 선도국과의 기술 격차 해소를 위해 중요하다고 인정한 분야임에도 불구하고 현재 우리나라에서는 관련 연구가 많이 이루어지지 않는 경우이다. 빅코드와 계산커널 오토튜너가 여기에 해당된다. 이들 연구들은 주로 개인연구 형태로는 지원되고 있는데, 국가적인 차원에서 중장기 대형과제로 확대하거나 관련 연구과제들의 연계를 강화할 필요가 있다.

본 연구에서 제시한 예비 후보 선발과 후보 검토 프로세스는 정부의 소프트웨어 과제 기획 프로세스에 접목할 수 있을 것으로 기대된다. 일반적으로 수요조사가 2사분기에 이루어지는데, 그 전에 기획팀에서 기술 선도국의 사업이나 과제를 대상으로 예비 후보를 선발하고, 나아가 후보 선정까지 이루어진다면 수요조사에서 제안된 것들과 함께 검토하여 과제 기획의 품질을 제고할 수 있을 것이다. 또한, 기술 선도국과의 기술 격차가 존재하는 상황에서 기술 선도국에서 최첨단의 추진 방향, 그것도 구체적인 기술 목표를 설정한 연구과제들을 종합적으로 검토함으로써, 지금보다는 더 도전적인 연구 목표를 설정하고 기술발전의 궤적을 고려하여 추진방향을 기획할 수 있다

SUMMARY

This study aims to find a way to promote the quantum jump strategy to drastically reduce the software technology gap with developed countries. In order to pursue the quantum jump strategy, it is necessary to use a method different from the existing one in the entire R&D process, from finding a task to conducting a business. For this purpose, this study focuses on developing improvement methods throughout the R&D process and developing methodologies for finding candidates for mid and long term large-scale SW R&D projects.

To explore the new R&D promotion system, the US DARPA program was examined and compared with the current government R&D planning and implementation system. The direction of the R&D promotion system for quantum jump is as follows.

① (Task planning based on real demand) DARPA projects business with a focus on connecting the actual demand of specific armies with the progress of science and technology. On the other hand, the domestic task does not place the research goal in actual use, so there is a weak organic link between the sub-tasks and difficulty in collecting the capability.

② (Strengthen top-down project planning) In DARPA, PM plays a leading role and researchers concentrate on exploring solution alternatives as technical experts. In Korea, researchers propose not only the search for resolution confrontation but also the use and demand of technology.

③ (Constant task planning and management) At present, the task planning of information and communication fields including domestic software is centered on the task planning committee. The task planning committee usually operates within a limited period of time, usually one to two months at a specific time, mainly at the end of the year. Despite a short period of time, candidates are quickly identified, confirmed, compared, selected, and detailed. It is therefore difficult to overcome the technological gap, especially the movement of the technology leadership.

④ (Re-establishing role of CP and task leader) In Korea, R&D task planning for software is centered on CP (Creative Planner) belonging to Information Technology Promotion Center (IITP). CP is in charge of project planning, but it is not involved in selection and management of project assignments, so it differs greatly from the authority and discretion of DARPA PM. CP is therefore responsible for DARPA's role as director, selecting and providing the highest level of staff to plan and carry out tasks. The PM, who is in charge of the PM task, needs a plan that does not carry out direct research like DARPA but concentrates on planning and running the task belonging to CP room.

⑤ (Ensuring the flexibility of project management) In order to minimize the uncertainty in the planning stage, it should support research research time and budget for about one year. In addition, the project manager should endeavor to be able to conduct street-level research by accepting all the methods such as contracts, public offerings, competition methods, etc. in accordance with the nature of the task. The task leader needs to constantly monitor strategic complementarities between the sub-tasks and ensure that the tasks can be consolidated and terminated at any time.

DARPA, NSF, EU, and others, we have identified four candidate candidates that we believe are worth pursuing in order to reduce technological gap in Korea in the future through the four-step process. The purpose of this study is to narrow down the technological gap and to find the target in a specific task unit. In addition, these tasks have the merit of concentrating only the technical part when evaluating the business performance and the ripple effect in the country already.

First, we selected 3,322 SW-related candidates from DARPA, IARPA, NSF and EU's Horizon2020 program, which constituted the candidate pool. For the candidate pools, DARPA, IARPA, NSF, and Horizon2020 are divided into CPS(Cyber-Physical Systems), Big Data Analysis, Cloud Computing, HPC(High-Performance Computing) and Open Source Software A total of 12 groups were selected for preliminary selection. A total of 108 tasks were selected as candidates for selection, with 10 or so selected for each group.

The preliminary candidates also gathered opinions on whether they would be suitable for mid- to long-term large-scale software R&D project planning by forming a professional

group of three or four persons in 12 areas. And to express only whether it is both extreme of recommendation and disposition hope. Basically, at least two experts were actively recommended and at the same time one candidate was selected as a candidate who did not wish to be excluded, and a total of 21 tasks became candidates for final evaluation.

The finalists were composed of 10 candidate review committees and 21 candidates were conducted through a three-point comparative review as shown in the previous table. The selection of the final candidates was made through a pair comparison between the candidates. A pairwise comparison is a comparison method that determines the preference of one of two pairs of objects. However, in the case of AHP and other pairs, the number of pairs increases exponentially as the number of candidates increases. Thus, we divided into three stages and used the short form pair method developed in this study.

In the first, four tasks were selected by combining DARPA and IARPA candidates. In the second, four tasks selected from the first and NSF candidates were selected. In the third, we compared Horizon2020 candidates to four tasks in the second. At this time, the final candidate should not only be superior to the competition candidates but also be absolutely worth pursuing in the mid- to long-term large-scale tasks in our country. Therefore, we selected one of mid- to long-term large-scale projects that are currently being compared in Korea and compared them. Of the 12 candidates, four were selected as finalists.

The final selected mid to long-term large-scale candidates are DARPA AIDA(Active Interpretation of Disparate Alternative), IARPA DIVA(Deep Intermodal Video Analytics) NSF BONSAI(An Open Software Infrastructure for Parallel Autotuning of Computational Kernels), and H2020 CROSSMINER(Developer-Centric Knowledge Mining from Large Open-Source Software Repositories).

The four finalists are divided into two major categories. The first group is in the mid- to long-term large-scale tasks currently under way in Korea, but the technical leadership has set new challenging goals. This includes AIDA and DIVA. AIDA is associated with the ExoBrain, intelligence information flagship project, and AI strategic strategy project. While these existing mid- to long-term large-scale projects are targeting only one of multi-source or multi-modal, the research projects of the technical leadership are expanding and combining multi-source, multi-modal and multi-hypotheses. Therefore, it is

necessary to reinforce the cooperation of researchers by newly planning tasks that can cover domestic tasks. DIVA also sets challenging research goals than the mid to long-term Deep View project. While the existing tasks mainly assume a single camera environment, the goal is to detect dangerous behavior in real time in a multi camera environment. Therefore, it is necessary to consider planning the project as a research project specializing in the protection of public facilities as well as the existing Deepview project and newly started challenging individual studies.

The second group is the area where experts are recognized as important for resolving the technological gap with technology leading countries, but there are not many related studies in Korea at present. This is the BONSAI and CROSSMINER. These studies are mainly supported by individual research forms, and it is necessary to extend them to mid to long term large-scale projects from the national level or to strengthen the linkage of related research projects.

The candidate selection and candidate review process proposed in this study is expected to be able to be applied to the government software project planning process. In general, demand surveys will take place in the second quarter. Before that, the planning team will be able to select prospective candidates for the projects and tasks of the technology-leading country and review them together with the proposals in the demand survey to improve the quality of the project planning. In addition, in the presence of technological gaps with technology-leading countries, it is necessary to establish research goals that are more challenging than the present, and to set the trajectory of technological development. The direction of the project can be planned.

CONTENTS

Chapter 1. Introduction

Chapter 2. R&D promotion system for quantum jump

Chapter 3. Software R&D Trend

Chapter 4. Methodology for finding R&D projects

Chapter 5. Finding candidates for SW R&D project

Chapter 6. Planning Direction of Final Candidates

Chapter 7. Conclusion

제1장 서론

제1절 연구 배경

소프트웨어 기술과 인재의 중요성이 증대되고 있는 가운데 우리나라의 소프트웨어 기술 수준은 선진국들과 격차를 보이고 있다. 정보통신진흥센터(IITP)의 우리나라의 기술수준을 조사한 결과에 따르면 2016년 기준으로 소프트웨어 분야는 기술적으로 가장 앞선 국가에 비해 70%대로 1.8년 내지 1.9년 뒤쳐져 있다. 2015년과 비교해서는 기술 격차가 조금 개선되었지만 컴퓨팅 시스템, 빅데이터 플랫폼, 응용 소프트웨어와 같이 세부 분야별로는 기술 격차가 악화된 경우도 있었다. 실제 상용 소프트웨어 전체로 볼 때 글로벌 시장을 선도하는 제품을 찾기 어려운 현실이다.

〈표 1〉 우리나라 소프트웨어 분야 기술수준

분류		기술격차(년)		기술수준(%)	
대분류	중분류	2016	2015	2016	2015
기반SW ·컴퓨팅	전체	1.8	2.0	71.7	74.1
	컴퓨팅 시스템	2.1	2.0	75.5	76.6
	High Performance컴퓨팅	-	2.8	-	66.3
	인공지능·인지컴퓨팅	2.2	2.4	73.9	70.5
	클라우드 컴퓨팅	1.6	1.8	72.4	77.0
	모바일 컴퓨팅	-	1.7	-	77.9
	빅데이터 플랫폼	1.7	1.6	78.2	76.3
SW	전체	1.9	1.9	76.7	76.8
	시스템 SW (OS, DBMS 등)	2.1	2.3	74.6	72.7
	미들웨어	1.9	2.0	76.9	76.3
	응용 소프트웨어	1.8	1.3	78.7	81.5

출처: IITP, ICT 기술수준조사

컴퓨터 과학(Computer Science) 연구자 커뮤니티에서의 리더십을 보더라도 우리나라는 열세이다. 연구자의 논문 편수와 인용도로 영향력을 평가하는 H-Index가 40 이상인 컴퓨터 과학자의 수를 보면, 미국이 993명으로 가장 많고, 영국이 그 뒤를 이어 125명이다. 3위는 독일로 104명이다. 4위 이후로는 캐나다(92명), 프랑스(71명), 스위스(56명), 네덜란드(53명), 이탈리아(48명), 오스트레일리아(48명), 이스라엘(35명)의 순서이다. 11위 이후로는 홍콩(29명), 스페인(26명), 일본(24명), 벨기에(23명), 중국(20명), 싱가포르(20명), 스웨덴(18명), 오스트리아(16명), 브라질(15명), 그리스(13명)의 순서이며, 21위 이후로는 타이완(12명), 인도(12명), 핀란드(9명), 대한민국(8명), 터키(7명), 아일랜드(6명)의 룩셈부르크(5명), 포르투갈(5명), 멕시코(5명), 폴란드(5명)로, 우리나라는 24위이다.¹⁾

소프트웨어의 기술 격차는 여러 가지 문제를 초래하고 있다. 그 중에서도 사회 전반에 걸쳐 적극적으로 소프트웨어를 도입하는데 걸림돌이 되고 있는 것이 가장 큰 문제라고 볼 수 있다. 소프트웨어의 가치는 활용을 통해 창출되는데, 상용 제품의 대부분이 미국을 비롯한 외산 제품이어서 소프트웨어 도입은 외화 유출을 의미하는 것으로 받아들여진다. 그리고 이런 인식은 소프트웨어 제품 도입에 소극적인 문화로 연결된다. 결과적으로 소프트웨어 산업의 성장을 가로막는다. 더욱 더 심각한 것은 제조업을 포함한 우리나라 주력 산업이 적극적인 소프트웨어 도입을 통해 경쟁력을 강화할 수 있는데, 그 기회를 활용하지 못하게 한다는 것이다.

따라서 소프트웨어 기술 격차를 획기적으로 해소하는 것은 소프트웨어 산업의 성장뿐만 아니라 우리나라 산업 전반의 경쟁력 향상을 이루는데 필요한 주요 과제 중 하나라고 할 수 있다. 즉, 소프트웨어 기술 격차를 줄이기 위한 국가 차원의 전략적 노력이 필요하다. 이는 새로운 소프트웨어 중심 시대를 대비하기 위해 피할 수 없는 선택이기도 하다.

우리나라 소프트웨어 기술 수준을 획기적으로 향상시키기 위해서는 단순히 기술격차 축소를 위한 추격형 혁신전략으로는 한계가 있다. 기술 선도국의 기술도 정체되어 있는 것이 아니라 향상되고 있기 때문에 기술 선도국과의 기술 격차는 일종의 무빙 타겟(moving target)이다. 기술 선도국의 기술 향상 속도보다 우리나라의 기술 향상 속도가 더 빨라야 한다. 이를 위해서는 과거 CDMA 기술개발로 단숨에 이동통신 선도국으로 도약했듯이 소프트웨어에서도 ‘퀀텀 점프(Quantum Jump)’ 전략이 필요하다.

퀀텀 점프 전략을 추진 방식은 다양하다. 단기간에 재원을 집중하는 형태로 추진될

1) <http://www.guide2research.com/scientists/by-country>

수 있다. 목표가 뚜렷하고 기술 방식에 선택의 여지가 없을 경우에 이런 방식이 적합하다. 반대로 목표나 기술 방식에 대한 불확실성이 높은 경우, 다수의 소규모 과제를 장기간 지원해서 원하는 목표를 달성할 수도 있다. 이 경우, 연구팀의 열정과 역량에 의해 성과가 크게 좌우될 가능성이 높다. 그러나 기술 수준의 획기적 발전을 이루기 위해서는 단기간의 집중 지원이나 장기간의 다수 대상의 소규모 지원의 장점을 모두 살린 중장기 대형 연구과제의 지원이 필요하다.

현재 우리나라에서도 국내 역량은 낮으나 세계시장 성장률이 높은 분야를 대상으로 국가혁신형(SW그랜드챌린지 프로젝트) 사업을 진행하고 있다. 엑소브레인(2013년), 딥뷰(2014년), 빅데이터 처리용 고성능 컴퓨팅(2015년), 브레인 컴퓨터 인터페이스(2017년) 등 매년 1개의 대형과제를 선정하고 있다. 또한 알파고 열풍이 있었던 2016년에는 국가혁신형 과제와 별도로 인공지능 분야에서 지능정보 플래그십 프로젝트(2016)와 인공지능 국가전략 프로젝트(2017)가 중장기 대형 SW 과제로 선정됐다.²⁾

하지만 이들 과제는 규모는 대형이기는 하지만 세부과제 간의 유기적인 연계가 미흡하여 과거 CDMA, DRAM 사례와 같이 역량을 결집시켜 국내 기술을 퀀텀점프 시키는 데 한계가 있다.

제2절 연구 목적 및 범위

본 연구의 목적은 선진국과의 소프트웨어 기술 격차를 획기적으로 줄이기 위한 퀀텀점프 전략의 추진 방안을 찾아보는데 있다. 퀀텀점프 전략을 추진하기 위해서는 과제 발굴에서 사업수행까지 R&D 프로세스 전체 과정에서 기존과는 다른 방법을 사용해야 한다. 이를 위해 본 연구에서는 R&D 프로세스 전반에 걸친 개선점 도출과 중장기 대형 SW R&D 과제 후보 발굴을 위한 방법론을 개발하는데 중점을 뒀다.

세부적으로 2장에서는 새로운 R&D 추진체계를 모색하기 위해 미국 DARPA 프로그램을 살펴보고 현재 정부 R&D 기획 및 수행체계와 비교 분석했다. 1958년부터 시작된 DARPA 프로그램은 통상적인 기술개발이 아닌 한계돌파(breakthrough)와 근본적인 개념의 변화를 가져오는 혁신적인 아이디어에 대한 지원을 통해 인터넷, GPS, 무인자동차, 인공지능 비서 등 새로운 제품과 시장을 탄생시켰다. 이와 같은 DARPA 프로그램의 성격은 본 연구에서 찾고자 하는 퀀텀점프 전략 추진방안이 나아가야 하는 방향을 제시해 줄 수 있을 것이다.

2) 상세내용은 <부록 2, 3> 참조

3장에서는 과제 후보 발굴에 앞서 소프트웨어 기술 선도 국가에서 생각하고 있는 소프트웨어 기술의 발전 방향 및 중점 영역에 대해 살펴보았다. 또한 소프트웨어 관련 연구소, 학회 등에서 발표된 자료를 분석하여 소프트웨어 전체를 관통하는 트렌드와 최근 각광 받고 있는 주제들을 살펴보았다.

4장에서는 기존 유망기술발굴 방법론의 문제점을 살펴보고 체계적인 과제 발굴을 위해 단계적으로 후보를 압축한 후 여러 기술 영역에 걸친 후보를 쌍대 비교하여 최종 후보를 선정하는 방법론을 제시하고 있다.

5장은 미국과 유럽연합에서 현재 수행중인 과제를 대상으로 본 연구에서 개발한 방법론을 사용하여, 향후 우리나라에서 획기적인 기술격차 축소를 위해 추진할만하다고 판단되는 과제 후보를 발굴했다. 이때 검토 범위를 기술 선도국의 과제로 한정된 것은 본 연구의 목적이 기술격차 축소에 있기 때문에 구체적인 과제 단위에서 목표점을 찾기 위함이다. 또한 이들 과제들은 이미 해당 국가에서 사업성, 파급효과 등을 검증하였기 때문에 기술적인 부분만 집중할 수 있는 장점이 있다.

6장에서는 발굴된 과제 후보에 대해 향후 과제 기획 방향을 제시하고 있다. 본 연구에서 발굴한 과제 후보는 현재 기술 선도국에서 진행하고 있는 과제이기 때문에 국내 여건과는 다를 수 있고, 실제 상세 기획에는 많은 시간과 인력이 필요하다. 그렇기 때문에 본 연구에서는 발굴된 후보에 대한 상세적인 추진방안 대신 향후에 과제로 추진할 때 고려해야하는 방향만을 제시한다.

제2장 퀀텀점프를 위한 R&D 추진체계

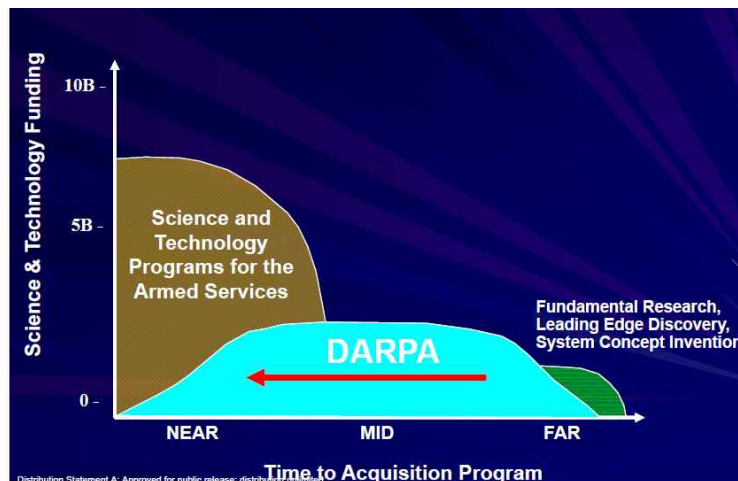
제1절 미국 DARPA 사례분석

1. 사업기획

DARPA는 도전적인 연구개발의 대명사인데, 중장기 대형 연구개발을 지향한다. 1958년 설립 이후 지금까지 수없이 많은 기술적 퀀텀 점프의 성과를 달성해 왔다. 인터넷의 전신인 ARPANET에서부터 위성항법장치(GPS), 스텔스 전투기, 터치스크린(touch screen), 애플의 시리(Siri) 등 다양하다. 현재 미국의 여러 부처는 DARPA의 성공모델을 받아들여 ARPA-E(에너지성), HS-ARPA(국토안보부), IARPA(CIA, FBI 등 정보관련 부처)와 같이 각자의 영역에서 혁신적이고 도전적인 기술을 개발하고 있다.

DARPA는 기관의 역할을 현존하는 최첨단 과학기술 지식을 신속하게 응용하여 군사적 수요와의 간극을 메우는 것으로 정의하고 있다. <그림 1>처럼, 각 군의 기존 연구조직들은 각 군의 단기적 수요나 기술적 필요에 대처한 단기적 연구개발에 치중되어 있어(좌측의 Armed Service S&T 영역) 새로운 과학적 지식, 혁신적인 신개념, 근본적인 발견/발명 등 장기적 연구개발 활동의 산물(우측의 작은 영역)과는 동떨어져 있음을 알 수 있다. 이에 DARPA는 스스로의 역할을 우측 영역에서 새로운 연구자와 아이디어를 찾아내어 좌측 영역의 연구원들이 신속하게 응용할 수 있게 하는 중기적 가교 역할(중간 영역)에 두고 있다.

<그림 1> 과학기술 발전에서의 DARPA 역할



이러한 DARPA의 역할에 따라 사업 선정을 하는데 있어 다음과 같은 3가지 기준을 따르고 있다.

〈표 2〉 DARPA의 사업 선정 기준

- 고위험 고성과(High Risk, High Payoff): 위험이 크지만 성공했을 경우의 성과가 매우 큰 사업
- 혁신적인 아이디어(Innovative Ideas): 통상적인 기술개발이 아니라 Breakthrough와 근본적인 개념사의 변화를 이끌 수 있는 혁신적인 아이디어를 구현할 수 있는 프로그램
- 가교기술(Bridge the gap between long-term basic research and short-term R&D): 장기적인 기초연구와 단기의 개발 수요 간의 간극을 연결해 줄 수 있는 중장기의 기반기술 연구 프로그램

출처: 송치웅 외(2013), 창의적인 연구개발을 위한 K-ARPA 시스템 구축방안, STEPI

DARPA는 현재가 아닌 미래의 지휘관이 무엇을 원할 것인가를 반문하고, 이에 따라 완전히 새로운 핵심 국방능력을 제공하는데 초점을 두고 있다. 이러한 역할은 성공적일 경우 획기적인 성과(High-Payoff)를 가져다 줄 수 있는 반면, 실패의 위험 또한 아주 높아(High-Risk), ‘고위험-고성과(High-Risk and High-Payoff)’ 성격의 연구개발 활동을 지향하는 것이다.

DARPA의 사업선정에 있어서 또 하나의 중요한 선정 기준은 최고의 아이디어와 최고 인재의 존재 여부이다. DARPA는 특정 기술 분야에 최고의 아이디어를 가진 최고의 인재를 영입할 수 없다면 절대 사업을 시작하지 않는 것을 철칙으로 하고 있다. 이들은 학계, 산업계, 연구소 등 각 분야로부터 기술적 우수성과 혁신성(Entrepreneurial)을 기준으로 DARPA의 핵심인 PM(Program Manager)으로 선발된다.

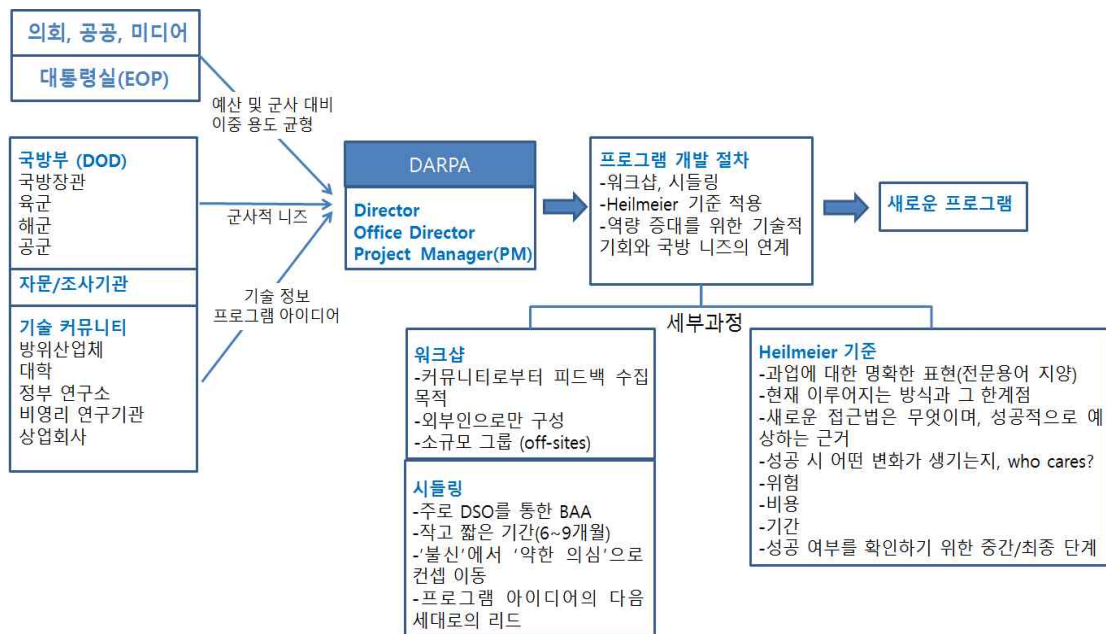
DARPA의 신규 사업 기획은 PM 주도 하에 아래 그림과 같은 프로세스로 진행된다. 사업 아이디어는 국방성, 기술 커뮤니티(방위사업체, 대학, 정부 연구소 등으로 구성) 등 외부기관뿐 아니라 동료 PM 등 다양한 채널에 의해 도출된다. 도출된 아이디어를 가지고 PM은 국방성 등 국방 분야 전문가들과의 수없이 많은 회의를 통해 그들의 수요를 거듭 확인하고, 그 수요를 충족하는데 걸림돌이 되는 기술적 문제를 정의한다.

그리고 기술 전문가인 PM 자신과 DARPA의 동료 PM, 그리고 외부 전문가 커뮤니티를 협조를 얻어 기술적 해결 가능성을 검토한다. 이런 내부 검토 외에도 공개 워크숍이나 시들링(seedling)이라고 하는 탐색 연구 등을 이용하여 불확실성을 줄이려는 노력

도 함께 하여 불확실성과 위험요인을 관리할 수 있도록 사업을 기획한다.

한편, PM은 워크샵, 시들링을 통해 나온 결과를 반영하여 아래 그림에 나와 있는 Heilmeier 기준에 의거하여 사업기획서를 작성한다. 사업 기획 안은 Director, 동료 PM 등의 내부 동료 검토(peer review)를 거쳐 사업 개시 여부가 최종 결정된다.

〈그림 2〉 DARPA 신규사업 기획 프로세스



2. 조직 및 사업 운영의 특징

① 위험을 감수하는 작고 민첩한 조직

DARPA는 고위험-고성과(High Risk, High Pay-off) 과제를 다루기 위해 위험을 회피하지 않고 실패를 감수하는 조직 문화를 가졌다. 민첩하고 유연한 연구를 추진할 수 있도록 자체 연구 설비를 보유하지 않는 대신 외부의 자원을 활용하고 이를 위해 PM의 활동을 집중 지원한다. PM은 각 과제의 연구를 책임지고 주도한다.

복잡한 보고 단계로 인해 연구가 지연될 소지가 큰 관료 체제와 달리 조직의 수장인 Director, Office Director, PM의 간략한 구조로 되어 있고 Office Director의 주요 업무도 적합한 PM을 발탁하고 지원하는 것이다.

DARPA의 구성원은 총 300여 명 수준인데, 과제 진행에 전념하는 프로그램 매니저가 150여 명 내외이고 나머지는 PM을 위해 각 프로그램의 제반 지원 업무를 담당한다.

비록 DARPA가 국방부 소속이지만 Director, PM 등 핵심 인력은 대부분 민간인이며 현역 군인 또는 군 출신은 10% 내외에 불과하다. 이들은 대학, 민간 연구소, 기업 또는 군에서 연구(Research) 경력을 갖춘 교수, 연구원, 엔지니어 출신의 석·박사 학위 소지자들이다. 상당수는 벤처 기업을 경영한 경험도 가지고 있어 프로그램의 진행, 신기술이 적용된 제품 개발, 사업화에 대해서도 밝다.

② PM의 강한 리더쉽

DARPA는 연구의 창의성과 혁신성을 최대한 보장하기 위해 PM에게 폭 넓은 재량권을 부여하고 있다. 연구 과정에서 정부 부처나 DARPA 등 조직 내외의 간섭에 구애받지 않도록 PM이 독자적인 결정권을 보유하게 한 것이다. 그래서 저명한 학자 등 외부 전문가들의 평가에 의존하는 NIH와 같은 여타 기관들과 달리 DARPA는 연구 과제 의 최종 선정에서부터 프로젝트 팀 구성과 연구 진행, 관련 예산의 집행에 이르기까지 프로그램의 진행 전반에 걸쳐 PM이 개입, 결정하도록 하고 있다. 이렇게 폭넓은 권한 때문에 DARPA의 PM은 각 프로그램의 Virtual CEO라고 불리기도 한다.

PM은 평균 4~6년 임기의 계약직인데, 연구 과제에 충실할 수 있도록 프로그램 기간에는 근무의 안정성을 보장받는다. 프로그램이 종료되면 DARPA에서 경험한 혁신적 제품의 사업화 역량을 인정받아 대학, 연구소, 기업 및 군으로 복귀하기도 하고 DARPA의 연구 결과를 사업화해서 창업하기도 한다.

③ 유연한 사업 운영

DARPA의 단위 사업(program)은 평균 4년 동안 10~40백만 달러의 예산으로, 한 두 개의 대학을 포함한 5~10개 내외의 연구조직과의 계약 형태로 수행된다. 연구 수행기관은 앞서 설명했듯이 외부 전문가의 평가가 아닌 PM이 전적으로 결정한다. 사업 운영방식도 PM이 정하는데 자율 주행 자동차 대회, 재난 대응용 로봇 개발 대회 등 기술 경연 대회와 같은 특별한 형태의 연구 프로젝트를 진행하기도 한다.

대부분의 경우 새로운 사업은 새로운 PM의 선임과 동시에 기획하여 그의 임기와 함께 종료되는 것을 원칙으로 하고 있다. 즉 새롭게 선출된 매니저는 취임과 동시에 새로운 아이디어와 혁신적 기술과 함께 DARPA에 입성한다. 새로운 아이디어는 새로운 사업의 형성을 의미한다. 대규모 주요 기술개발은 훨씬 긴 시간을 소요하지만, 이런 경우 일련의 단계적인 소규모 프로그램으로 나누고, 각 단계별로 명확한 목표를 세워

서 일련의 프로젝트를 수행하게 된다. 또한 각 프로젝트는 일단 종료되면 그것으로 끝이며 이어지는 후속 프로젝트는 없는 것을 원칙으로 한다. 같은 연구 분야에 같은 PM이 새로운 프로젝트를 시작하면 외부에는 종료된 프로젝트의 연속으로 비칠 수도 있지만, DARPA로서는 이 프로젝트는 완전히 새로운 결정이며 과거 프로젝트의 존재 유무는 전혀 고려의 대상이 아니다.

제2절 R&D 추진체계 개선방향

1. 사업기획

① 실질 수요에 기반한 과제기획

앞 절에서 살펴봤듯이 DARPA는 구체적인 실질 수요와 과학기술의 혁신적 진보를 연결하는데 초점을 두고 사업을 기획한다. 과거 우리나라에서 퀀텀점프 전략의 대표적 성공사례인 CDMA의 경우도 마찬가지로 아직 상용화 가능성이 검증되지 않은 기술을 기반으로 디지털 방식의 이동통신 망을 구축한다는 구체적인 수요 달성을 목표로 추진됐다. 하지만 1장에서 언급한 국내 SW 분야 중장기 대형 과제는 연구개발 완료 후에 그 결과를 실제 사용하는 것을 목적으로 하고 있지 않아 세부 과제간 유기적 연계가 약하고 역량 결집이 어려운 문제점을 가지고 있다.

물론 통상 문제 등으로 과거와 같이 상업적 제품 개발에 정부에서 지원하기는 어렵지만 국방, 재난재해 등과 같이 사회적 문제 해결 영역은 가능하다. 그렇게 때문에 혁신적인 최신 SW기술을 활용하여 사회 문제를 실질적으로 해결할 수 있는 상용화 제품을 개발하는 R&D 과제를 발굴하고 추진해야 한다. 또한 연구개발 결과를 정부가 수요자가 돼서 시범사업과 조달 등을 통해 기술을 검증하고 민간에 이전시키는 노력이 필요하다.

② 하향식(Top-Down) 과제 기획 강화

과제 발굴 프로세스 또한 DARPA와 큰 차이가 있다. DARPA에서는 PM이 주도적인 역할을 하고 연구자들은 기술 전문가로서 해결 대안 탐색에 집중하는 반해, 우리나라에서는 연구자들이 해결 대결의 탐색뿐만 아니라 기술의 사용처와 그 수요까지 함께 제안한다. 즉, PM의 역할이 제한적이고, 수요자가 거의 관여하지 않은 상황에서 과제

기획이 이루어지는 것이다. 과제에서 상정한 수요는 가상 수요여서 연구개발에 성공했을 때 실제 사용될 가능성은 매우 낮다.

또, 우리나라의 과제 기획은 상향식 방식을 중심으로 이루어진다. 대표적으로 기술수요조사가 있다. 기술수요조사는 1개월 내외의 제한된 기간에 이루어지는데 문서를 통해 진행된다. 심지어 DARPA와 동일한 국방 분야에서도 정형화된 문서 형식을 활용하여 상향식으로 아이디어가 모아진다. 그런데 우리나라의 기술수요조사에는 원천적으로 결함을 가지고 있다. 기술수요조사에서 제안된 것들을 기초로 하기 때문에 기술수요조사의 품질에 크게 좌우된다. 물론 DARPA의 방식에서도 상향식 과제 제안 절차가 있다. 그러나 이 절차는 두 가지 면에서 차이를 보인다. 하나는 기간이 1년 이상으로 매우 길어 사실상 상시 의견을 받는다는 것이다. 다른 하나는 외부로부터 제안 받은 의견은 참고할 뿐 실질적인 의사결정은 DARPA의 PM에 의해 수정되고 결정된다는 것이다.

③ 상시적 과제 기획 운영

현재 국내 소프트웨어를 비롯한 정보통신 분야의 과제 기획은 과제기획위원회를 중심으로 이루어진다. 과제기획위원회는 통상 특정 시점, 주로 연말에 1~2달 정도의 제한된 기간 내에 운영된다. 짧은 기간에도 불구하고 후보 발굴과 확정, 비교 검토, 선정, 상세 기획까지 빠르게 진행된다. 그러다 보니 기술 격차 극복을 위해, 특히 기술 선도국의 움직임을 포괄적으로 검토하기 어렵다. 상향식으로 제안한 과제 후보나 기획 위원회의 논의를 통해 선정된 주제와 관련된 기술에 대한 선도국의 움직임만 선별적으로 확인하는 경향이 있다.

이런 문제를 해결하기 위해서는 일명 상시 기획이 이루어져야 한다. 연말의 특정 시점에 과제 기획 활동이 집중되지 않고 연중 과제 기획을 위한 활동이 진행되는 것이다. 예를 들면, 과제기획위원회를 구성하기 전부터 기술 선도국에서 새롭게 시작하는 과제들을 분석할 수 있다. 과제 정보를 체계적으로 수집하고 전문가 그룹에 기술 격차 측면에서 해당 과제의 영향력을 검토할 수 있다. 그리고 그런 활동을 통해 우리나라의 과제 기획에서 검토해야 할 기술 선도국의 과제 목록을 작성해서, 하향식 기술 기획의 기초자료로 활용할 수 있다.

2. 사업 운영 방식

① CP와 과제책임자의 역할 재정립

우리나라에서 소프트웨어 관련 R&D 과제기획은 정보통신진흥센터(IITP)에 소속된 CP(Creative Planner)를 중심으로 이루어진다. 민간 전문가인 CP는 선발과정을 거쳐 2~5년 정도 계약 또는 파견직으로 근무한다. CP가 과제기획을 하지만 과제책임자 선발 및 운영에는 관여하고 있지 않아 DARPA PM의 권한 및 재량과는 크게 차이가 난다. 과제 기획과 운영의 분리는 공정성 제고의 효과를 가져올 수 있지만, 퀀텀점프 전략에서 추구하는 창의·도전적 연구에 필요한 추진력과 유연성을 확보하는데 한계가 있다.

한편, 실질적인 성과 창출을 위해서는 과제 총괄 책임자에게 DARPA의 PM과 같은 권한과 재량권을 부여하는 방법이 있을 수 있다. 하지만 DARPA PM의 강한 리더십은 이해관계의 중립성으로부터 나온다. DARPA의 경우 직접 연구를 수행하지 않으면서 사업 책임을 맡고 있기 때문에 우수 연구진 확보가 사업 성패를 크게 좌우한다. 그래서 DARPA PM들은 연구진 확보에 총력을 쏟을 수밖에 없다. 반면 국내 R&D 과제의 총괄 책임자는 직접 연구도 수행하기 때문에 이해관계에서 자유롭지 못한 상황이다.

그러므로 퀀텀점프를 위한 R&D 사업의 경우 DARPA의 벤치마킹하여 추진체계를 새롭게 구성할 필요가 있다. 우선 CP는 DARPA의 Director 역할을 담당하여 과제를 기획하고 추진할 최고 수준의 인력을 선발하고 지원하는 업무를 담당한다. PM 역할을 담당할 총괄 과제책임자는 DARPA처럼 직접 연구를 수행하지 않고, CP실에 소속되어 과제 기획과 운영에만 전념하도록 하는 방안이 필요하다.

② 과제 운영의 유연성 확보

퀀텀점프를 위한 R&D과제는 기본적으로 고위험-고성과를 지향하기 때문에 기존과는 다른 과제 운영 방식이 필요하다. 기획단계에서 불확실성을 최소화할 수 있도록 1년 정도의 탐색 연구 시간 및 예산을 지원해야 한다. 또한, 과제 성격에 맞게 과제책임자가 단독 수의 계약, 공모, 경쟁 방식 등 모든 방식을 용인하여 현장 중심(Street-level)의 연구 수행이 될 수 있도록 노력하여야 한다. 과제책임자는 세부 과제 간의 전략적 보완관계를 끊임없이 모니터링하고 언제든지 과제의 통합 및 종료가능하도록 하는

것이 필요하다.

이러한 유연한 프로젝트의 수행을 위해서는 국회로부터 Grant 방식의 불특정 예산(Earmark free)을 받도록 노력하여야 할 것이며, 이월 예산을 용인하고 이월에 따른 예산 삭감이 없도록 보장해야 한다.

무엇보다 근본적이고 파괴적인 혁신을 추구하기 때문에, 실패를 폭넓게 용인하고 수용할 수 있는 문화를 정착시키는데 전력을 다하여야 한다. 이를 위해서는 기존 국가연구개발사업에서 적용하는 평가위원회를 통한 선정, 진도 및 사후 평가를 지양하고, 실질적인 사업평가(Program review)를 활성화하여 사업 오류 방지, 실패로 부터의 교훈, 후속 사업과의 연계 등에 실질적으로 도움이 되는 평가를 이끌어 내는데 중점을 두어야 한다. 다만 부정행위 및 도덕적 해이 등을 방지하기 위한 철저한 회계감사는 필요하다.

제3장 소프트웨어 R&D 트렌드

제1절 기술 선도국

1. 미국

미국 연방정부의 ICT 분야 연구개발 총괄 프레임워크인 NITRD(Networking and Information Technology Research and Development) 사업에서 2011년 12월에 ‘Future of Software Engineering Research’ 라는 제목의 보고서를 발표했는데, NITRD의 핵심 분야 중 하나인 SDP(Software Design and Productivity)의 미래 투자 방향을 제시한 보고서이다. 동 보고서는 2010년에 ACM SIGSOFT 개최한 제18회 소프트웨어 엔지니어링 기반에 관한 국제 심포지엄(FSE: International Symposium on the Foundation of Software Engineering)에 포함된 Foundations of Software Engineering and Software Design and Productivity (FSE/SDP) Workshop on the Future of Software Engineering Research (FoSER)에서 발표되고 논의된 내용을 정리한 것이다. 이 보고서에서 다룬 다섯 가지 주제는 다음과 같다.

- 소프트웨어 개발 및 이용 지원 (Help people produce and use software-intensive systems)
- 미래의 복잡한 시스템 설계 (Design complex systems for the future)
- 신뢰할 수 있는 소프트웨어 집약적 시스템 개발 (Create dependable software-intensive systems)
- 소프트웨어 개발에 관한 의사결정 향상 (Improve decision-making, evolutions, and economics)
- 학문적 발전 및 연구방법 향상 (Advancing our discipline and research methodology)

① 소프트웨어 개발 및 이용 지원

사람들의 소프트웨어 개발과 이용을 향상시키기 위해 소프트웨어 생명주기(lifecycle) 전반에 관여되어 있는 개인과 그룹을 서로 연결하는 방법을 제시했다. 지금까지 무료 또는 오픈 소스 소프트웨어(FOSS)에서 메일링 리스트나 리포지토리(repository)가 유용하게 사용되어 왔는데, 이를 더욱 확대하고 더 좋은 협력 도구를 만들고 도입할 필요가 있다. 특히, 소프트웨어 엔지니어 사이의 연결을 확대하는 것은 소프트웨어 회사들의 채용 확대, 개발자간의 지식 공유, 그리고 소비자까지 포함한 개인들과의 협력을 확대할 것으로 기대된다.

구체적으로 개별 소프트웨어 의사소통 방식 개선, 소프트웨어 개발자들의 커뮤니티 형성 및 유지, 소프트웨어 개발 전문업체(ISV: Independent Software Vendor)와 분야별 활용 조직과의 연결, 운영체제 중심의 생태계 형성, 응용 분야별 소프트웨어 스택 형성 등 다양한 것들을 필요로 한다. 이를 위해 개발자들의 연대에 대한 인식 변화와 소프트웨어 프로젝트 관련 정보에 대한 접근성 향상, 소프트웨어 개발에 대한 분석 활성화가 필요하다.

또 사용자 중심의 프로그래밍을 위한 소프트웨어 엔지니어링 연구에 투자해야 한다. 소프트웨어 생산과 이용에 사용자 참여를 확대할 필요하며, 이를 위해 사용자들의 실제 사용 방법과 환경에 대한 정보가 개발자 커뮤니티에서 충분히 제공될 수 있게 한 것이다. 여기에는 최종 사용자들의 이용 동기, 관심, 제반 능력, 전문 분야에 특화된 지식에 대한 이해도 포함된다.

사람의 개입이 많은 시스템에 대한 연구가 필요하다. 일명 사이버-소셜 컴퓨팅 시스템(Cyber-Social Computing Systems) 과학을 발전시켜야 한다. 사이버-소셜 컴퓨팅 시스템 과학이란 인간과 기계, 소프트웨어의 상호 작용과 협력에 관한 현실과 바람직한 방향을 탐구하는 것으로, 특히 사람들이 많이 개입하는 국방, 보건의료 분야에서의 정보 처리를 강조했다.

② 미래의 복잡 시스템(complex systems) 설계

보건의료, 국방, 안전, 도시 등 사회의 여러 분야에 시스템 도입이 확산되면서 복잡 시스템의 도입이 보편화되는 추세에 대비해야 한다. 독립적으로 개발된 여러 단위 시스템들이 복합 시스템을 구성하는 것이 일반화될 것이기 때문에 시스템의 시스템, 시스템 오케스트레이션(orchestration)이 중요한 이슈로 부상할 것이다. 이를 위해 무엇보다

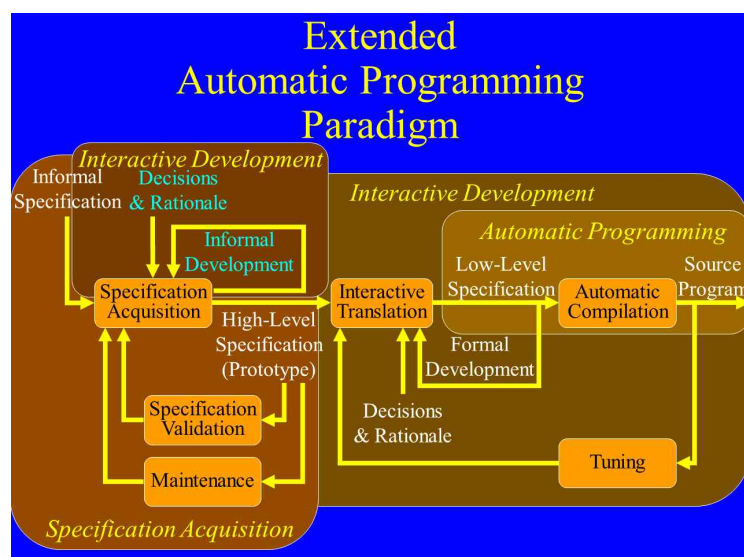
다 여러 분야를 아우르는 커뮤니티(multi-disciplinary communities)를 형성하는 것이 선행되어야 한다.

③ 신뢰할 수 있는 소프트웨어 집약 시스템(software-intensive) 개발

이 주제는 소프트웨어 개발의 생산성 향상과 소프트웨어 집약 시스템의 진화에 관한 것이고, 일명 ‘구성에 의한 수정(correct by construction)’의 패러다임이 적용하는 것이다. 일반적으로 소프트웨어 엔지니어링이란 요구사항(requirement)과 명세(specification) 변화에 대응해서 코드를 수정하는 것인데, 자동화된 프로그래밍에서는 기계의 도움을 얻어 프로그래밍 과정에서 사람의 개입을 최소화시키는 것을 목표로 한다.

프로그래밍의 자동화는 소프트웨어 엔지니어링의 과학적 근간을 흔드는 매우 중요한 주제이기도 하다. 아래 그림에서 보는 것처럼 이 패러다임이 성공하기 위해서는 요구사항과 명세 과정을 자동화해야 한다. 그런데 요구사항은 통상 의뢰인이 자연어로 개발자에게 표현한다. 따라서 컴퓨터는 의뢰인이 자연어로 표현한 요구사항을 별도의 입력 없이 그 자체로 인식할 수 있어야 프로그래밍 자동화가 가능하다. 그리고 인지한 요구사항을 바탕으로 컴퓨터 스스로 소프트웨어를 설계하기 위해서는 소프트웨어 디자인의 구성요소에 관한 오픈소스 기반의 말뭉치(corpus)가 체계화되어 있어야 한다. 마지막으로 설계된 소프트웨어의 품질을 측정하고 보장할 수 있어야 하는데, 특히 자동화되어있어야 한다.

<그림 3> 자동화된 프로그래밍 개념도



조금 더 구체적으로 살펴보면, 요구사항과 명세를 표현하는 방식을 발전시켜야 한다. 이를 위해 요구사항 정의와 명세 작성에서 해결해야 할 도전적 문제를 찾고 요구사항을 명세로 전환하는데 필요한 범용 프레임워크를 개발해야 한다. 그리고 이것들을 달성하기 위해 소위 명세 언어 서밋(Specification Languages Summit)의 개최를 제안했다.

한편, 현재 프로그래밍 언어와 같이 매우 제한된 수단으로 컴퓨터와 소통하는 방식에서 벗어날 필요가 있다. 인간이 사용하는 자연어로 명령을 내려 소프트웨어 프로그래밍이 가능하게 만드는 것이다. 가능한 대안으로 UML이나 ProcedureSpace와 같은 시스템을 예로 들고 있다. 한편 동 보고서에서는 향후 현재 소프트웨어 개발자에게 요구사항을 설명하는 것과 동일한 방식으로 컴퓨터에게 직접 요구사항을 설명할 날이 올 것을 기대된다고 말하고 있다.

소프트웨어가 탑재된 시스템이 장기간 사용되는 경우가 늘어나면서 해당 시스템의 변화 대응이 중요한 이슈로 부각되고 있는데, 그 시스템에 탑재되어 있는 소프트웨어 또한 마찬가지이다. 플러그인(plug-in) 기능을 지원하는 IDE의 사용이 확산되는 것도 이런 추세를 반영하는 것이다. 따라서 진화, 즉 변화를 인식하고 대응할 수 있는 소프트웨어 개발이 필요하다. 소프트웨어의 진화를 명시적으로 고려하고 있는 프로그래밍 모델이 필요한데, 예를 들면 CoDPOM(Context-Driven Process Orchestration Method)가 그 중 하나이다.

또, 소프트웨어의 신뢰성(dependability)을 평가할 수 있어야 한다. 이를 위해 시스템의 구성요소는 각각 독립적으로 분리되어 독립된 분석이 가능해야 하며, 소프트웨어 분석 기법이 소프트웨어의 작동 방식을 제대로 반영해야 한다. 이때 분석에 소요되는 비용이 있더라도 최소화하여야 해야 하는 요구조건이 있기 때문에 분석 결과를 여러 목적으로 재활용하는 것도 중요한 이슈 중 하나이다. 그리고 이는 궁극적으로 품질 소프트웨어 시스템(quality software systems)의 개발을 목표로 한다.

이때, 확장성이 큰 정적 분석(static analysis)과 수정된 부분만 분석하는 차등 분석(differential analysis)을 결합한 차등 정적 분석(differential statics analysis)이 좋은 대안이 될 것으로 평가했다. 즉, 기존의 소프트웨어 코드에는 결점이 없다고 가정하고 수정된 코드로 인해 발생할 수 있는 오류를 분석하는 것이다. 따라서 절대적으로 오류가 없음을 보장하는 것은 아니다. 그러나 시스템이 복잡해지고 대규모화 되는 지금의 추세에 적합한 대안이 될 것이다.

④ 소프트웨어 개발에 관한 의사결정 향상

소프트웨어 개발 과정에서 필연적으로 발생하는 자체 개발 여부, 출시 일자 등과 같은 의사결정을 개선해야 한다. 지금까지는 관련자들의 경험과 관행이 이런 의사결정에 절대적 영향을 미쳤지만 향후 소프트웨어 개발과정에서 발생하는 데이터가 늘어나고 수집되면서 데이터 분석의 필요성이 증가하고 있다. 예를 들면, 1,000만 라인 이상의 코드로 구성된 모질라(Mozilla) 웹브라우저의 개발에는 약 20만 건의 커밋(commits)과 50만 건 이상의 버그 보고가 있었다.

⑤ 학문적 발전 및 연구방법 향상

소프트웨어 엔지니어링의 학문적 발전과 연구방법을 고도화하기 위해 연구문제의 종류를 체계화하여야 한다. 특히, 실증 분석에서 의미 있는 연구문제를 발굴하고 해당 연구문제에 대한 유용한 해결 대안을 제시해야 한다. 연구결과의 재현이 가능해야 하며, 연구 데이터의 공유가 이루어져야 한다. 또, 평가 항목을 체계화하여야 하는데 이를 위해서는 연방 정부의 재원이 투자되는 대규모 실험 연구가 필요하다. 연구방법 고도화를 위해서는 현재 개발과 테스트가 분리되어 있고, 각각 소프트웨어 개발 과정에서 많은 시간이 소요되는 문제를 해결해야 한다. 그 대안 중 하나로 개발 도구에 코드 분석 기능을 내장하는 것이 될 수 있다. 그리고 소프트웨어 엔지니어링의 핵심은 개발자 등^외 사람의 인식과 행동이기 때문에 사회과학자와의 협력을 장려해야 한다.

2. EU

2016년 11월에 유럽연합은 “Research Priorities in the area of Software Technologies” 라는 제목의 보고서를 발표했다. 이 보고서는 IEEE Computer Society 2022 보고서, EU의 Knowledge Future 보고서, 세계경제포럼(WEF)의 Deep Shift 보고서 등 여러 자료와 2014년과 2016년에 개최된 소프트웨어 엔지니어링 국제학회(ICSE: International Conference on Software Engineering)의 소프트웨어 엔지니어링의 미래(FOSE: Future of Software Engineering) 트랙에서 발표된 논문들을 참고하여 정리한 것이다. 보고서의 구성은 소프트웨어 개발에 영향을 주는 과학기술 트렌드와 응용도메인별 소프트웨어의 영향, 소프트웨어 엔지니어링의 도전과제, 국가적 현황의 순서로 되어 있다.

소프트웨어 개발에 영향을 주는 과학기술 트렌드로는 8가지를 제시했다. 소프트웨어로 정의되는 만물(SDX: Software Defined Anything) 및 코드로서의 인프라(IAC: Infrastructure as Code)와 클라우드 컴퓨팅(cloud computing), 빅데이터 및 분석(big data and analytics), 범용 메모리 컴퓨팅(universal memory computing), 멀티코어 아키텍처(multi-core architecture), 양자 컴퓨팅(quantum computing), 자연스런 사용자 인터페이스, 기계학습을 제시했다.

한편 소프트웨어가 영향력을 크게 발휘할 8개 응용 영역을 제시했는데, 자율주행차, 공개 지적재산권(open intellectual property), 대규모 공개 온라인 교육(MOOC: Massive Open Online Course), 사물인터넷(IoT: Internet of Things), 3D 프린팅, 생명과학, 금융 기술, 인더스트리 4.0이 그것이다.

소프트웨어 엔지니어링 도전과제로 8가지를 제시했다. 첫째, 확장성과 복잡성을 해결해야 한다. 개별 시스템의 증가뿐만 아니라 시스템과 시스템의 연결이 늘어나면서 일명 시스템의 시스템(systems of systems)이 보편화 될 것으로 전망된다. 결과적으로 소프트웨어 개발은 복잡성을 해소하는 방향으로 이루어져야 한다. 현재보다 더 유연한 방식으로 단위시스템들이 연결이 되어야 한다. 또한 애자일 개발(agile development)이 보편적으로 도입되고 놀라운 확장성을 가져야 한다. 조직 내 개발팀이나 타 개발팀과의 협업뿐만 아니라 조직 밖까지 확대될 수 있어야 한다.

둘째, 소프트웨어 개발 도구와 추상화에서 발전이 있어야 한다. 특히, 소프트웨어가 일상생활에 광범위하게 영향을 미치면서 필요한 기술의 개발을 전문 개발자들에게 맡기는 것은 충분하지 않다. 따라서 전문가 외에 사용 범위를 넓힐 수 있는 프로그래밍 도구의 개발이 절실하다. 여기에는 프레임워크, API, 방법 등의 발전도 포함한다. 같은 이유로 소프트웨어 개발의 저변을 확대할 수 있는 MOOC에 대한 투자도 필요하다.

셋째, 보안과 프라이버시, 신뢰성을 확보할 수 있어야 한다. 이들 세 가지 속성은 소프트웨어가 탑재된 시스템의 사회적 확산을 결정짓는 중요한 요소여서 향후 소프트웨어 집약형 시스템의 성패가 여기에 달려 있다. 이 목표는 시스템 아키텍처, 프로그래밍 언어, 디자인 방법, 개발 도구 등 다양한 영역에서 종합적으로 향상되어야 달성할 수 있는데, 설계 단계나 개발 도구 자체에 이런 기능을 사전에 포함시키는 것이 한 방법이다. 예를 들면, 시스템 설계 방법론에 보안, 프라이버시, 신뢰성을 확보할 수 있는 형태로 발전시키는 것이다.

넷째, 소프트웨어 자체에 대한 분석과 그로 인한 성능 향상이 필요하다. 소프트웨어 개발이나 구동으로 인한 매우 많은 데이터가 생성된다는 점을 볼 때, 빅데이터 분석이 이들과 무관한 것이 신기할 정도이다. 소프트웨어 개발과 구동 과정의 데이터 분석을 통해 신뢰성, 사용성, 개발 효율, 수익성을 개선시킬 수 있을 것이다.

다섯째, 극단적으로 많은 수가 참여하는 협업 개발을 가능하게 해야 한다. 지금까지 인터넷 혁명으로 인해 소프트웨어 패키지 리포지토리(repositories) 또는 포지(forges)와 가 등장했고, 그로 인해 소스코드 재활용에서 큰 진전을 이루었다. 그리고 현재 오픈 소스 소프트웨어의 경우, 수천에서 수만 명이 한 모듈을 함께 개발하고 있다. 이런 추세는 앞으로도 지속될 것이다. 따라서 더 많은 수의 협업 개발이 가능할 수 있게 소프트웨어 개발 방법을 발전시킬 필요가 있다.

여섯째, 소프트웨어 개발 프로세스의 발전이 이루어져야 한다. 1980년대까지는 프로세스 단계별 관리가 엄격하여 순차적 개발이 보편화되어 있었는데, 1990년대 이후 애자일 개발 방식이 도입되었고 생산성이 향상되었다. 그런데 최근에는 요구사항 변경에 대한 실시간 또는 자동화된 개발 및 조정의 필요성이 대두되고 있다. 다음으로는 소프트웨어의 위험 요인과 보상 요인을 수치화하여 관리하고 조정하는 방법에 대한 필요성이 증가하고 있다. 그 다음으로는 테스트의 중요성이 증가하고 있으며, 테스트 방법의 향상이 필요하다.

일곱째, 소프트웨어 개발과 운영의 통합이다. IT 인프라가 클라우드 기반으로 변모하면서 개발과 운영이 동일한 환경에서 가능하게 되었고, 그 결과 최근 DevOps이 각광을 받고 있다. 특히, 소프트웨어 개발에 비해 소프트웨어 운영은 여전히 과학적 발전을 이루지 못한 분야인 점을 감안하면 개발과 운영의 통합은 소프트웨어 개발 생산성 향상을 위한 새로운 원천이 될 것이다.

여덟째, 환경 측면에서의 지속가능성을 확보하는 것이다. 즉, 에너지 효율적인 시스템 개발에 소프트웨어가 기여해야 한다. 최근 고성능컴퓨팅(HPC) 분야에서도 소프트웨어를 통한 에너지 효율 달성이 관심 주제가 되고 있는데, 이런 추세가 반영된 것으로 볼 수 있다.

제2절 연구자 커뮤니티

1. IEEE

정보통신 분야를 대표하는 전문가 집단인 전기전자기술자협회(IEEE: Institute of Electrical and Electronics Engineers)는 전기공학과 전자공학, 통신공학, 컴퓨터 공학과 그 응용 분야에 몸담고 있는 약 42만 명의 연구자가 회원으로 참여하고 있으며, 연간 예산이 4억 1,300만 달러인 거대 조직이다.

IEEE의 미래연구방향위원회(Future Directions Committee)에서는 협회와 관련된 새로운 연구 분야를 이니셔티브(initiative) 형태로 선정해서 관련 이벤트를 개최하고 있다. 2017년 말 기준으로 현재 진행 중인 이니셔티브로는 10개가 있다. 빅데이터(Big Data)와 브레인(Brain), 사이버 보안(Cybersecurity), 디지털 센스(Digital Senses)³⁾, 그린정보통신(GreenICT), 리부팅 컴퓨팅(Rebooting Computing)⁴⁾, 스마트 물질(Smart Materials), SDN(Software Defined Networks), 5G가 그것이다.

종료된 주제로는 클라우드 컴퓨팅(Cloud Computing)과 생명과학(Life Sciences)⁵⁾, 교통 전기화(Transportation Electrification), 사물인터넷(Internet of Things), 스마트시티(Smart Cities)의 6개가 있다. 이 중 사물인터넷과 스마트시티는 2017년에 종료되었다.

한편, 새롭게 제안된 이니셔티브 후보로는 21개가 있다. 3D 이미징과 멀티미디어(3D Imaging and Multi-Media)와 3D 프린팅(3-D Printing), 자율시스템과 사회적 영향(Autonomous Systems and their Societal Impact), 바이오연료(Bio Fuel), 블록체인(Blockchain), 탄소 격리(Carbon Sequestration), 깨끗한 물(Clean Water), 코그니티브 라디오(Cognitive Radio), 디지털 문화유산과 스마트 환경(Digital Cultural Heritage and Smart Environments), 환경공학(Environmental Engineering), 포그 컴퓨팅과 네트워킹(Fog Computing and Networking), 인메모리 컴퓨팅(In Memory Computing), 공간 인터넷(Internet of Space)⁶⁾, 마이크로 유체역학(Micro Fluidics), 나노포어 유전자 시퀀싱(Nanopore DNA Sequencing), 재부팅이 필요 없는 새로운 운영체제(NewOS: RebootlessOS), 모바일 시스템에서의 전원(Power in Mobile System), 신뢰성(Reliability), 스웜 로봇틱스(Robotics Also Swarms), 차세대 안전 중요 시스템(Safety Critical

3) 다양한 감각의 인식과 재현, 합성 등을 연구

4) 점진적인 발전과 혁명적인 전환 등 모든 가능성을 둔 새로운 형태의 컴퓨팅 기술 연구

5) IEEE 회원들과 생명공학의 전문가들의 협력 필요성에 대한 인식 제고를 목적으로 함

6) 장소 중심으로 인터넷 서비스를 제공하는 네트워크 체계에 관한 연구

Systems, Next Gen), 공생 머신(Symbiotic Machines)이다. 이 중에서 블록체인은 2017년에 새롭게 추가된 주제이다.

〈표 3〉 IEEE Future Direction Initiatives

구분	Initiative	비고
종료 (6개)	Cloud Computing	
	Internet of Things	2017년 종료
	Life Sciences	
	Smart Cities	2017년 종료
	Smart Grid	
	Transportation Electrification	
진행 (10개)	Big Data	
	Brain	
	Cybersecurity	프라이버시 이슈 포함
	Digital Senses	시각/청각/촉각/후각/미각 등
	Green ICT	
	Rebooting Computing	디바이스에서부터 UI까지
	Smart Materials	
	Software Defined Networks (SDN)	
	Symbiotic Autonomous Systems (SAS)	인간과 자율시스템의 공존
	5G	
신규 후보 (21개)	3D Imaging and Multi-Media	3D 센싱/전송/시각화
	3-D Printing	
	Autonomous Systems and their Societal Impact	현재 진행 중인 SAS와 연계
	Bio Fuel	
	Blockchain	2017년 신규
	NewOS: RebootlessOS	
	Power in Mobile Systems	
	Reliability	
	Robotics Also Swarms	
	Safety Critical Systems, Next Gen	
	Symbiotic Machines	
	In Memory Computing	
	Internet of Space	
	Micro Fluidics	
	NanoPore DNA Sequencing	
	Carbon Sequestration	
	Clean Water	
	Cognitive Radio	
	Digital Cultural Heritage and Smart Environments	
	Environmental Engineering	
	Fog Computing and Networking	

한편, IEEE에서는 다수의 저널을 발간하고 있고, 컴퓨터 과학 연구자 커뮤니티에 큰 영향을 미치고 있다. 그 중에서도 컴퓨터 과학 전반의 다루는 Proceedings of the IEEE(2016년 IF: 9.237)와 컴퓨팅 분야의 유망기술을 다루는 IEEE Transactions on Emerging Topics in Computing(2016년 IF: 3.826), 소프트웨어에 특화된 IEEE Software(2016년 IF: 2.190)의 특집호 주제를 수집하여 분석하였다.

Proceedings of the IEEE의 최근 특집호 주제⁷⁾는 다음과 같다. 폰노이만 구조와 CMOS를 탈피하고자 하는 차세대 컴퓨팅 아키텍처뿐만 아니라 CPS의 디자인 자동화와 안전성, 소프트웨어 공학 등이다. 이들 주제는 2017년과 2018년에 논문 접수가 마감되고 2018년 특집호에 게재될 예정이다.

<표 4> 2018년도 Proceedings of the IEEE 특집호 주제

- | |
|--|
| <ul style="list-style-type: none"> • Non-Silicon, Non-von Neumann Computing • Tactile Internet • Silicon Photonics • Rethinking PCA for Modern Datasets: Theory, Algorithms, and Applications • Applications of Graph Theory • Smart Cities • Design Automation for Cyber-Physical Systems • Machine Ethics: the design and governance of ethical AI and autonomous systems • Foundations and Trends in Localization Technologies • From High Level Specification to High Performance Code • Safe and Secure Cyber-Physical Systems • Small Satellites |
|--|

IEEE Transactions on Emerging Topics in Computing(TETC)은 기존 IEEE Transaction 시리즈에 등장하지 않는 새로운 주제를 전문적으로 다루는 저널이다. TETC의 특집호 주제⁸⁾로는 Scholarly Big Data와 Design of Reversible Computing Systems이 있다.

IEEE Software는 소프트웨어 전반에 대한 이슈를 다루는 잡지이다. 2017년 12월 이

7) <http://proceedingsoftheieee.ieee.org/preview-upcoming-issues/>

8) <https://www.computer.org/web/tetc>

후인 출간 예정인 특집호 주제들은 다음과 같은데, 소프트웨어 엔지니어링에 관한 것들이 많다.

〈표 5〉 IEEE Software 특집호 주제

주제	게재 시기
<ul style="list-style-type: none"> • Software for Context-Aware and Smart Healthcare • Contextual Variability Modeling 	2017년 11/12월호
<ul style="list-style-type: none"> • Software Safety and Security Risk Mitigation in Cyber-physical Systems • Actionable Analytics in Software Engineering 	2018년 2월호
<ul style="list-style-type: none"> • Release Engineering 3.0 • Microservices 	2018년 4월호
<ul style="list-style-type: none"> • Managing Software Platforms and Ecosystems • Collaborative Modeling in Software Engineering 	2018년 8월호
<ul style="list-style-type: none"> • 50 Years of Software Engineering 	2018년 9/10월호
<ul style="list-style-type: none"> • The Social Developer: The Future of Software Development • Cloudware Engineering 	2018년 11/12월호
<ul style="list-style-type: none"> • Large-Scale Agile Development • Building Long-Lived Adaptive Systems 	2019년 3/4월호

2. ACM⁹⁾

컴퓨팅기기학회(ACM)는 1947년에 설립되어 약 10만 명의 회원을 보유하고 있다. ACM도 IEEE와 마찬가지로 다수의 저널의 발간하고 있고, 컴퓨터 과학 연구자 커뮤니티에 큰 영향을 미치고 있다. 그 중에서도 컴퓨팅 분야의 유망기술을 다루는 Journal on Emerging Technologies in Computing (JETC) (2016년 IF: 1.367)가 있는데, 시스템 아키텍처와 하드웨어와 소프트웨어의 디자인과 통합을 주요 주제로 다룬다. 이 저널의 2015년 이후 특집호 주제는 다음과 같다.¹⁰⁾

9) <http://jetc.acm.org/announcements.cfm>

10) <http://jetc.acm.org/announcements.cfm>

〈표 6〉 ACM Journal on Emerging Technologies in Computing 특집호 주제

- Nanoelectronic Circuit and System Design Methods for Mobile Computing Era
- Alternative Computing Systems
- Hardware and Algorithms for Learning On-A-Chip
- Frontiers of Hardware and Algorithms for On-chip Learning
- Silicon Photonics
- Emerging Networks-on-Chip: Designs, Technologies, and Applications
- Neuromorphic Computing: Architectures, Models and Applications

3. 기타 저널의 특집호(Special Issues) 주제

IEEE와 ACM가 아닌 다른 기관에서 발간하고 있는 컴퓨터 과학 분야의 저널들이 있는데 이들 저널의 특집호 주제 중 소프트웨어와 관련된 것들이 있다. 상세 내용은 〈부록 6〉을 통해 확인할 수 있다.

여러 저널 중에서 ELSEVIER에서 발간하는 Future Generation Computer Systems (2016년 IF: 3.997)에서 여러 특집호 주제를 제안했다. 구체적인 목록은 다음과 같다. 클라우드와 엣지 컴퓨팅, 블록체인이 1회 이상 등장했다.

〈표 7〉 Future Generation Computer 특집호 주제

- Accountability and Privacy Issues in Blockchain and Cryptocurrency
- Cyber Threat Intelligence and Analytics
- Autonomous Cloud
- High Performance Services Computing and Internet Technologies
- Edge of the Cloud
- Novel edge computing-inspired approaches and paradigms for mobile IoT applications
- Blockchain and Decentralization for Internet of Things
- Computation Intelligence for Energy Internet
- The Digitization of the Individual
- Smart Data in Future Internet Technologies and Cloud Computing
- Technological Innovations in Digital Transformation

그 외에도 특집호 논문을 모집하는 여러 저널들이 있다. 예를 들면, 컴퓨터과학 분야의 대표 저널인 Pattern Recognition(2016년 IF: 4.582)에서는 “Deep Learning for Computer Aided Cancer Detection and Diagnosis with Medical Imaging”의 특집호를 준비 중에 있다.

Pattern Recognition Letters(2016년 IF: 1.995)에서는 아래와 같은 특집호를 준비 중에 있다. 빅데이터와 딥러닝, 시각 지능에 관한 것들로 감시시스템이나 헬스케어 등 응용 분야 주제도 함께 선정되어 있다.

〈표 8〉 Future Generation Computer 특집호 주제

- Graphonomics for e-citizens: e-health, e-society, e-education
- Learning Compact Representations for Scalable Visual Recognition and Retrieval
- Multiple-task Learning for Big Data (ML4BD)
- Learning and Recognition for Assistive Computer Vision
- Virtual Pattern Recognition Techniques for Non Verbal Human Behavior (NVHB)
- DLVA: Advances in Deep Learning and Visual Analytics for Intelligent Surveillance Systems

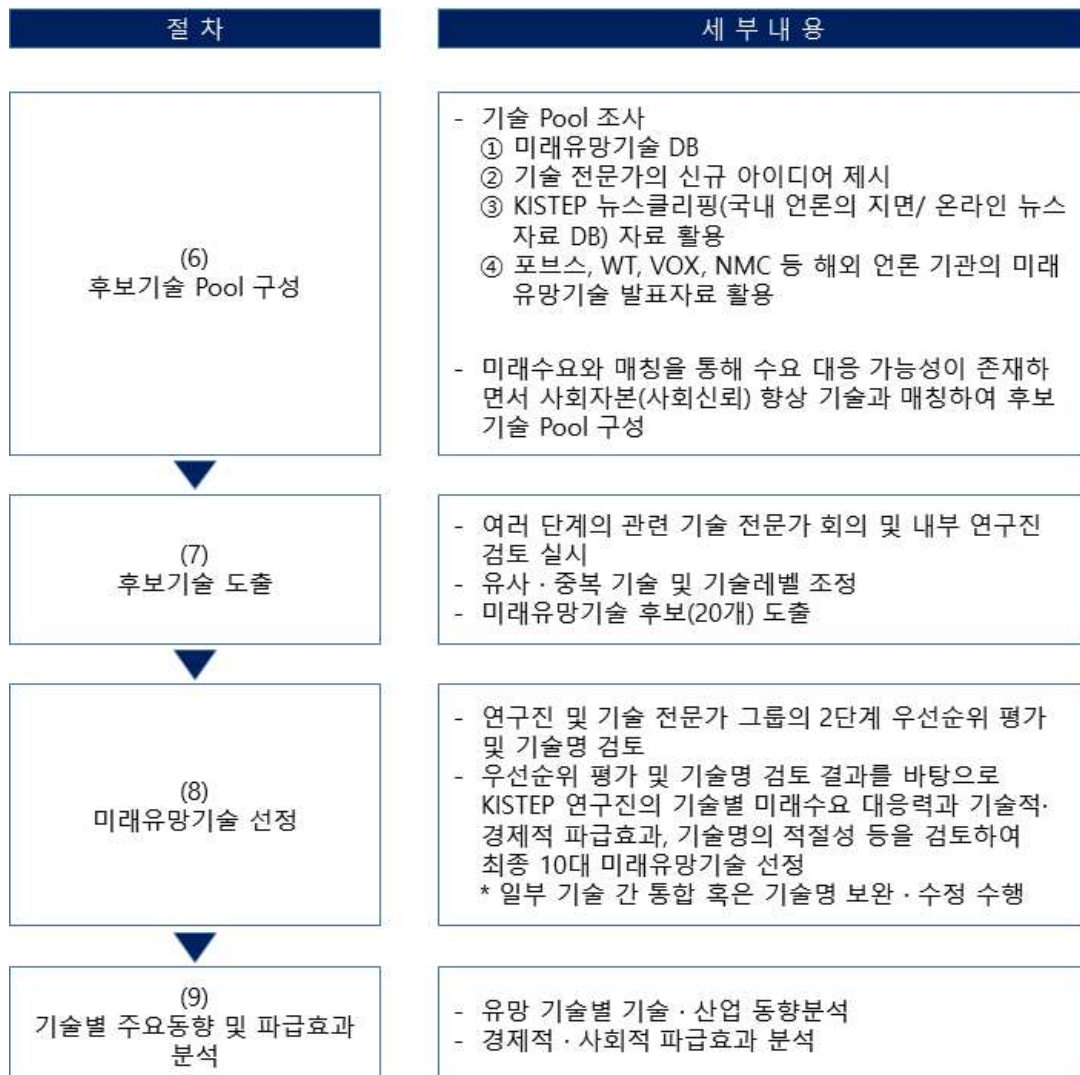
Information Systems Journal(2016년 IF: 4.122)에서는 “Indigenous Theory”와 “The Digitization of the Individual”, “Cybercrime and Computer ‘Digital’ Forensics”을 주제로, WILEY의 Software: Practice and Experienced(2016년 IF: 1.025)에서는 “클라우드 컴퓨팅과 사물인터넷, 빅데이터 분석의 통합(Integration of Cloud, IoT and Big Data Analytics)”을 주제로 논문을 모집하고 있다.

제4장 중장기 대형 R&D과제 발굴 방법론

제1절 기존 유망기술 선정 방법론

하향식 기획의 경우, 유망기술을 선정하여 상세기획을 추진하는 방식을 택하는 것이 일반적이다. 즉, 인공지능, 브레인-컴퓨터 인터페이스와 같은 분야를 먼저 선정하고 해당 분야에서 개별 과제를 기획하는 것이다. 매해 KISTEP, KISTI, IITP 등 국내 주요 기관에서 유망 기술 분야를 선정하여 발표하고 있다. 그 방식은 비슷하다. 아래 KISTEP의 사례에서 보는 것처럼 후보기술 Pool을 구성하고, 후보기술로 그 수를 압축하여 최종적으로 미래 유망기술을 선정한다.

〈그림 4〉 KISTEP의 유망기술 선정과정



그런데, 이런 방식에는 몇 가지 약점을 내포하고 있다. 첫째, 전략성이 결여된 채 후보기술 Pool을 구성하게 되면 후보기술 Pool을 최대한 많이 확보해야 한다. 여기에 기술 기획 활동의 대부분이 할애될 수밖에 없고, 이는 기획 과정의 효율성에 심각한 문제를 발생시킨다. 둘째, 후보기술 도출과정에서 후보기술 Pool을 통합·변형·배제 등의 작위적 조작이 가해질 수밖에 없다. 대체로 후보기술의 범위가 넓을수록 중요성이 부각되어 후보로 선정될 가능성이 높다. 또 최근에 언론 등에서 많이 회자되는 주제가 후보기술의 범위에 포함될수록 후보선정에 더 유리하다. 즉, 후보기술의 제목과 범위가 후보 선정에 절대적 영향을 미칠 수 있다. 셋째, 후보기술 내에서의 상대평가만으로는 절대 기준의 충족 여부를 확인할 수 없다. 즉, 후보기술 중 최선의 것을 선정할 수 있으나 후보기술 전체가 잘못 구성된 경우 적절하지 없는 선택이 될 수 있다. 넷째, 최종 선정은 전문가들의 평가로 결정되는 것이 일반적인데, 20개 내외 후보를 대상으로 3~10개의 평가 항목을 평가한다. 이때 후보기술의 수와 평가 항목의 수가 많아 평가 품질이 저하될 가능성을 항상 가지고 있다.

위에서 제기한 문제들을 해결하기 위해서는 다음과 같은 보완이 필요하다. 첫째, 후보기술 Pool 구성의 노력을 최소화하기 위해서는 모든 기술을 검토 대상으로 삼지 않고 사전에 전략적으로 분야를 선택을 하여 검토 범위를 제한해야 한다. 예를 들면, 본 연구에서와 같이 소프트웨어로 그 분야를 제한하고 중장기 대형 과제에 적합한 후보로 제한하면 후보기술 Pool 대상이 줄어들어 그것을 구성하는데 소요되는 노력이 많이 줄어들 수 있다. 둘째, 후보기술을 그 명칭과 범위가 명확하고 구체적인 것들로 구성해야 한다. 예를 들면, 현재 국내외에서 진행하고 있는 실제 과제를 대상으로 하면 후보기술의 명칭이나 범위의 모호성이 줄어들 것으로 기대된다. 구체적인 논문이나 특허도 가능하다. 셋째, 후보기술 내에서의 상대평가로 인한 문제를 해결하기 위해 상대평가와 함께 절대평가도 함께 도입해야 한다. 후보 선정에 사용되는 절대평가 방법은 다양하다. 그 중 하나가 일정 기준 점수 이상이 되지 못하면 상대평가에서 상위에 랭크되더라도 배제하는 방법이다. 또 다른 방법으로는 이미 통과한 실제 과제를 비교 대상을 삼는 것이다. 즉, 기존에 추진된 과제보다는 더 의미 있는 후보기술을 선정하는 것이다. 넷째, 최종 선정을 위한 후보의 수와 평가 항목의 수를 줄여서 전문가들의 평가 품질을 제고할 수 있다. 그리고 평가 항목과 후보의 수가 줄게 되면 후보들 간의 직접적인 상대 비교, 즉 쌍대평가도 가능하다.

제2절 본 연구의 과제 발굴 방법

1. 기본 방향

본 연구에서 사용한 과제 발굴 방법론은 5가지 기본 원칙을 가지고 개발됐다. 첫째, 기술 선도국과의 기술 격차 해소에 직접적으로 기여하는 것을 목적으로 하고 있기 때문에 기술 선도국에서 실제 수행하고 있는 과제를 검토 대상으로 했다. 1절에서 살펴봤듯이, 국내외 기관이 발표한 미래기술을 후보 Pool로 하는 일반적인 유망기술 도출 방법은 범위가 넓고 통폐합 과정에서 왜곡이 발생하는 등 여러 가지 문제를 가지고 있기 때문에 본 연구에서는 검토범위를 기술 선도국의 과제로 한정했다. 특히, 기술 선도국에서 최근 기획한 과제들을 검토함으로써 그들이 인식하고 있는 기술 발전 경로를 추정하여, 그들과의 기술 격차를 줄이기 위해 우리나라는 어떤 노력을 해야 하는지 시사점을 얻을 수 있다. 즉, 기술 선도국의 움직임을 체계적으로 분석함으로써 해당 기술의 발전 경로를 명확하게 이해하고, 후발 주자로서 기술 발전 단계를 생략하거나 새로운 경로를 창출할 수 있는 방안을 도출할 수 있다.

둘째, 기술 선도국 과제에 대한 검토와 선별은 기술적 가치 평가를 중심으로 수행했다. 현재 기술 선도국에서 수행하고 있는 과제들은 모두 일련의 기획 과정을 통과한 것들이다. 즉, 정부 재원이 투자할만한 가치 있다는 것이다. DARPA 사업의 경우에는 실질 수요와 과학적 진보까지 모두 포괄한다. 다시 말해서 검토 대상의 기술 선도국 과제들 중 상당수는 기술 수요가 검증되었다고 가정할 수 있다. 따라서 본 연구에서는 기술 선도국의 과제들에 대해 기술개발의 성공 가능성과 성공했을 때의 기술적 파급효과만을 기준으로 검토했다.

셋째, 중장기 대형 소프트웨어 R&D과제 발굴은 기술 선도국의 전체 과제 목록에서 예비 후보, 예비 후보에서 후보, 후보에서 최종 후보를 선정하는 형태, 즉, 수를 줄여나가는 방식으로 진행했다. 이때 과제에 대한 정보 수집의 노력과 전문가 검토 사이에 균형을 이루고 노력을 최소화하기 위해 전반부에는 연구팀이 많이 관여하고 후반부에는 전문가 그룹의 역할이 큰 형태이다. 즉, 본 연구의 연구팀이 정부의 과제 기획 프로세스 상의 CP실 역할을 대신한다고 가정하는 것이다.

넷째, 예비 후보에서 후보, 후보에서 최종 후보 선정의 과정은 복수의 전문가들 사이에 공통된 의견을 통해 선발하는 것을 원칙으로 했다. 만약 3명의 전문가가 검토하였다면 최소 2명 이상이 검토 대상으로 추천한 경우에만 선발한다. 또, 2명 이상의 전문

가가 추천하였더라도 단 한 명의 전문가라도 적극적으로 배제를 희망할 경우에는 선발하지 않았다.

다섯째, 최종 후보 선발에 상대적인 순위 뿐만 아니라 절대적인 우수성을 확인할 수 있도록 보완했다. 새롭게 선발한 최종 후보가 후보들 중 상대적으로 전문가들로부터 좋은 평가를 받은 것이지만, 그것만으로는 충분하지 않다. 그래서 우리나라에서 소프트웨어 중장기 대형 과제로 선정되어 현재 수행중인 과제와 한글과 같이 우리나라만의 특이한 주제에 관한 신규 기획 과제를 대상으로 하여 비교 평가도 실시했다.

2. 과제 발굴 프로세스

본 연구의 과제 발굴 프로세스는 아래 표와 같이 5개 단계로 진행됐다.

〈표 9〉 중장기 대형 SW과제 발굴 프로세스

구분		그룹	담당
1	후보 Pool 구성	미국: DARPP, IARPA, NSF 유럽연합: H2020	연구팀
2	예비 후보 선발	기관 전체: DARPA, IARPA 기술 분야 구분: NSF, H2020	연구팀
3	후보 선정	기관 전체: DARPA, IARPA 기술 분야 구분: NSF, H2020	전문가 평가 (분야당 3~4인)
4	최종 후보 선정	1차: DARPA + IARPA 2차: 1차 + NSF (단, OSS는 3차에 포함) 3차: 2차 + H2020 + 기존 과제	후보 검토 위원회 (10명)
5	향후 기획 방향	4개 최종 후보	연구팀

① 후보 Pool 구성

앞의 기본원칙에서 기술했듯이 소프트웨어 기술 선도국인 미국과 EU에서 현재 진행 중인 과제를 중심으로 검토 후보 Pool을 구성했다. 본 연구의 목적이 퀀텀점프를 위한 R&D과제를 발굴하는 것이기 때문에 이와 같은 특성을 보유한 미국의 DARPA와 IARPA(Intelligence Advanced Research Projects Activity)의 과제를 우선적으로 포함했

다. 또한 개발 연구보다는 기초연구에서 퀀텀점프 할 수 있는 기술이 나올 수 있기 때문에 미국의 기초연구를 담당하는 NSF의 과제를 후보 Pool에 포함시켰다. 사실 미국이 소프트웨어 분야에서 가장 기술 선도국이기 때문에 미국에서 진행 중인 과제들만 검토하는 것으로도 충분하다. 그러나 본 연구에서는 연구 주제의 누락을 방지하기 위해 EU의 대표적 R&D프로그램인 Horizon2020(이하 H2020로 표기)의 과제들을 함께 검토했다.

② 예비후보 선발

예비 후보 선발은 연구팀의 자체 분석을 통해 이루어졌다. 이때 예비 후보 선발은 기본적으로 기관별로 구분해서 이루어졌다. 즉, DARPA 과제는 DARPA 과제끼리, IARPA 과제는 IARPA 과제끼리 묶어서 그 중에서 각각 우리나라와 기술 선도국의 기술 격차에 해소에 기여할 가능성이 높은 과제를 선발하였다. 미국 NSF와 유럽연합의 H2020의 과제 수는 DARPA나 IARPA보다 매우 많았다. 그래서 기관별로 구분해서 예비 후보를 선발한 것뿐만 아니라 기관 내에서도 기술 분야별로 구분해서 예비 후보를 선발했다. 이때, 사용한 분야는 5개인데, CPS(Cyber-Physical Systems), 빅데이터 분석, 클라우드 컴퓨팅, HPC(High-Performance Computing), 오픈 소스 소프트웨어(OSS)다. 각 그룹은 10개 내외의 예비 후보를 선발하였다.

③ 후보 선정

후보 선정 역시 DARPA와 IARPA는 기관 전체 사업을 대상으로 이루어졌고 NSF와 H2020은 기관의 분야별로 이루어졌다. 즉, DARPA 전체 사업, IARPA 전체 사업, NSF의 5개 분야, H2020의 5개 분야 총 12개 그룹으로 나누어 후보 선정이 이루어졌다. 각 그룹은 3~4명의 전문가들로부터 중장기 대형 소프트웨어 R&D과제 기획 대상으로 적합한 여부의 의견을 수렴하였는데, 적극 추천과 배제 희망의 양 극단에 해당하는지만 표현하게 했다. 기본적으로 최소 2명 이상의 전문가가 적극적으로 추천하고 동시에 한 명이라도 배제를 희망하지 않은 것들을 후보로 선정하였다.

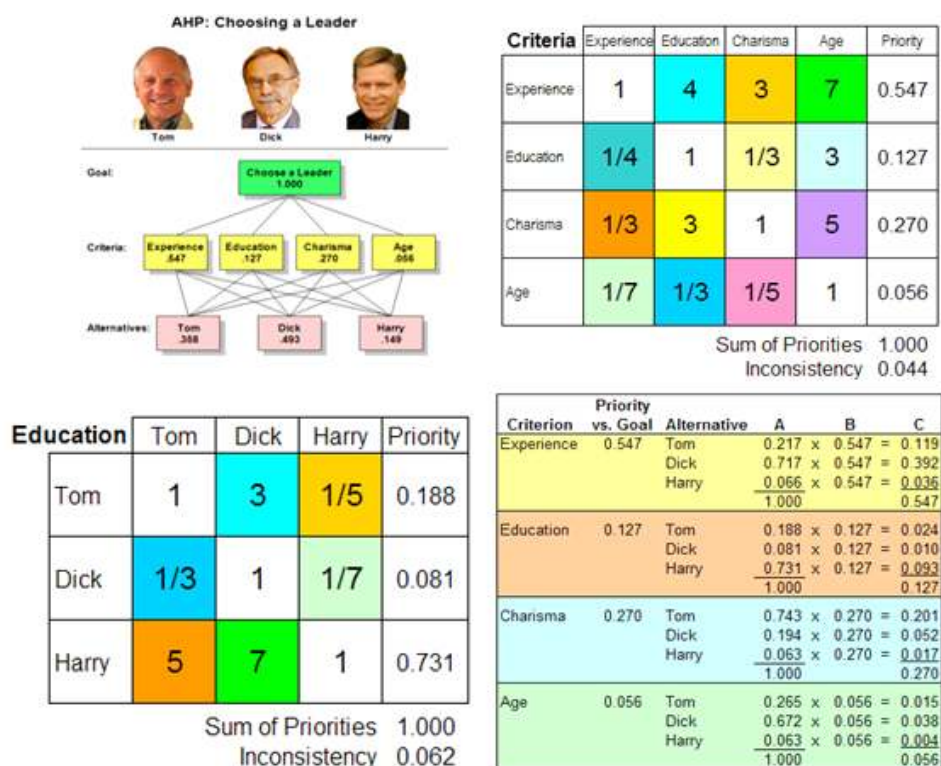
④ 최종후보 선정

최종 후보 선정은 12개 그룹으로부터 선발된 후보들을 3차에 걸친 비교 검토를 통해 진행하였다. 1차 비교 검토는 DARPA와 IARPA의 사업 중 선정된 5개 사업을 대상으로 하였고, 2차 검토는 1차 검토에서 생존한 후보와 NSF의 4개 분야, 즉 CPS와 빅데이터

분석. HPC, 클라우드 컴퓨팅 분야의 후보를 합하여 이루어졌다. 마지막 3차 검토는 2차 검토의 생존 후보와 NSF의 오픈 소스 소프트웨어(OSS)와 H2020의 5개 분야, 그리고 기존 중장기 대형 과제까지 모두 합하여 비교하였다. 그리고 마지막 3차 비교 검토를 마친 후 생존한 후보를 최종 후보로 선정하였다.

한편, 최종 후보 선정은 후보들 간의 쌍대 비교를 통해 이루어졌다. 쌍대비교란 비교 대상들을 짝을 지어 둘 중 하나의 선호를 결정하는 비교 방식이다. 대표적인 방법이 토마스 사티(Thomas L. Saaty)가 개발한 AHP(Analytic Hierarchy Process)가 있다. AHP는 통상 4개 이내의 후보들을 대상으로 계층화된 평가 항목을 이용하여 합리적인 의사결정을 하도록 도와준다. AHP의 적용 사례는 아래의 그림에서 확인할 수 있다. 3명의 후보 중 리더를 선택하는 문제에 적용한 것이다. 평가항목은 4개로 경험과 교육, 카리스마, 연령이다. 먼저 평가항목간의 쌍대비교를 통해 평가항목의 가중치를 결정한다. 아래의 예에서는 경험이 0.547로 가장 높고, 그 뒤를 이어 카리스마(0.270), 교육(0.127), 연령(0.056)의 순으로 결정되었다. 각 평가항목에 대해 3명의 후보를 쌍대비교하고, 각 평가항목의 가중치를 곱하고, 다시 후보별로 그 값을 합산하여 각 후보의 값을 계산한다. 아래의 예에서는 3명 후보의 값이 0.358과 0.493, 0.149로 계산되었다.

<그림 5> AHP의 적용 사례: 3명의 후보 중 리더 선택



출처: 위키피디아 영문판

그런데 AHP를 비롯한 쌍대비교에서는 후보의 수가 하나씩 증가할수록 쌍대비교의 개수가 기하급수적으로 증가하기 때문에 후보가 많은 경우 일반적으로 적용하지 않는다. 또, 평가항목의 수가 적거나 계층화된 구조를 갖지 않는 경우에 AHP를 사용하지 않아도 된다. 단순한 평가방식으로 충분하다. 이런 점을 감안하여 본 연구에서는 10개가 넘는 후보들을 대상으로 비교하고 평가항목도 2개로 매우 적어서 AHP를 그대로 적용하지 않고 약식의 쌍대비교를 사용하였다. 본 연구에서 사용한 약식 쌍대비교는 아래 그림에서 보는 것처럼 후보들간의 모든 쌍에 대해 쌍대비교를 하지 않고 후보 수보다 1개 더 많은 쌍대비교를 하게 설계되었다. 또, 일반적인 쌍대비교에서는 1배, 3배 5배, 7배, 9배 등 최대 9배까지를 쌍대비교의 값으로 사용하는데 반해, 기술의 비교에서는 통상 그 값의 차이가 크지 않은 점을 감안하여 최대 3.0배까지만 선택하도록 하였다.

<표 10> 본 연구의 약식 쌍대 비교의 응답지 양식과 응답결과 예시

더 높다	3.0 배	2.5 배	2.0 배	1.8 배	1.6 배	1.4 배	1.2 배	1.1 배	동 일	1.1 배	1.2 배	1.4 배	1.6 배	1.8 배	2.0 배	2.5 배	3.0 배	더 높다
후보1											○							후보2
후보2															○			후보3
후보3					○													후보4
후보4							○											후보5
후보5							○				□							후보1

위의 응답 예시를 기준으로 각 후보의 값을 계산하면 다음과 같다. 먼저 후보1은 1.0으로 가정한다. 후보2는 후보1보다 1.2배 더 높기 때문에 $1.20(=1.0 \times 1.2)$ 이 된다. 후보3은 후보2보다 2.0배 높아서 후보1의 $2.40(=1.2 \times 2.0)$ 배가 된다. 반면 후보3이 후보4보다 1.6배 높기 때문에 후보4는 후보1의 $1.50(=2.4/1.6)$ 배이다. 후보4도 후보5보다 1.2배 높기 때문에 후보5는 후보1의 $1.25(=1.5/1.2)$ 배이다. 마지막으로 후보5가 후보1보다 1.2배 높기 때문에 후보1은 $1.04(=1.25/1.2)$ 가 된다. 이때 처음에 가정한 후보1의 1.0과 쌍대비교를 통해 산출된 1.04 사이에 0.04의 차이가 발생하였다. 그러나 그 차이가 처음 가정한 1.0의 5% 미만이기 때문에 일관성에는 문제가 없다고 판단된다. 대신 후보1은 1.0과 1.04의 평균인 1.02를 부여받는다. 결과적으로 후보1은 1.02, 후보2는 1.20, 후보3은 2.40, 후보4는 1.50, 후보5는 1.25이다. 높은 값부터의 순위는 후보3와 후보4, 후보2, 후보5, 후보1이다.

그런데, 만약 마지막 쌍대비교를 동그라미(○) 대신에 네모(□)로 한다면 후보1의 계산값은 $1.5(=1.25 \times 1.2)$ 가 된다. 당초 가정된 1.0과는 0.5의 차이가 발생하여 50%의 오차가 발생한다. 평가자의 평가에 일관성이 없다고 할 수 있다. 이런 경우, 일관성이 없는 평가자의 평가를 배제하거나 평가자가 일관성을 갖추도록 안내하는 방법이 있는데, 본 연구에서는 후자를 선택하였다.

위의 예는 1개의 평가 항목에 대한 것이다. 따라서 각 평가 항목에 대한 쌍대비교의 값을 종합하여 전체 점수를 산정한다. 일반적으로 평가 항목의 중요도를 가중 평균한다.

3. 추진 체계

본 연구는 소프트웨어정책연구소와 한국전자통신연구원이 공동으로 연구팀을 구성하여 진행했다. 연구팀은 후보 출처 선정, 후보 출처로부터 예비 후보 선발, 최종 선정 후보의 추진방안 마련 등 본 연구 전반에 걸쳐 주도적 역할을 수행했다.

연구팀이 선발한 예비 후보들로부터 후보를 선정하는 과정은 각 분야의 전문가들에게 의뢰하였다. 기술분류체계의 중분류에 해당하는 분야별로 3~4인의 전문가에게 의뢰하여 다수가 추천하고 특별히 배제를 희망하지 않은 것으로 선정하였다.

후보검토 위원회는 외부 인사를 중심으로 구성했다. 총 10명의 위원 중에 8명이 외부 인사이고 2명은 연구팀이 속한 기관의 인사이다. 외부 인사로는 정부 연구개발사업의 기획·평가·관리 책임자 2명과 벤처투자전문가 2명, 학계 2명, 산업계 2명으로 구성했다. 이들 중에서는 전직 기획·평가·관리 책임자도 포함되어 있고, 벤처투자전문가 중에는 컴퓨터 공학 박사학위 소지자도 포함되어 있다. SW 분야가 광범위한 특성을 고려하여 분야를 넘나들며 전반적인 비교 평가를 할 수 있는 전문가들로 구성된 점이 특징이라고 할 수 있다.

제5장 중장기 대형 SW R&D 과제 발굴

제1절 후보 Pool 구성

1. DARPA

2017년 8월20일 기준으로 DARPA는 219개 과제를 진행하고 있다.¹¹⁾ DARPA의 6개 부서 중 소프트웨어와 가장 관련이 깊은 부서는 I2O(Information Innovation Office)인데, I2O는 46개 과제를 진행하고 있다.

DARPA 과제의 대표적인 특징으로는 단기간의 단일 과제로 끝나는 것이 아니라 복수의 지속적인 과제를 통해 투자하고 있다는 것이다. 예를 들면 이미지 인식, 동영상 이해 등의 시각지능 분야에서도 VIRAT(Video and Image Retrieval and Analysis Too), Mind's Eye, VMR(Visual Media Reasoning), MADCAT(Multilingual Automatic Document Classification, Analysis and Translation), VACE(Video Automatic Content Extraction) 등 오랜 기간 동안 과제를 추진해 왔다.¹²⁾ 언어 지능 분야에서도 GALE(Global Autonomous Language Exploitation), BOLT(Broad Operational Language Translation), DEFT(Deep Exploration and Filtering of Text), MEMEX 등의 과제에 투자해 왔다.¹³⁾ 또 하드웨어 연구를 핵심 연구 주제로 하는 MTO(Microsystems Technology Office)에서도 임베디드 소프트웨어 등 일부 관련 연구를 수행하고 있다.

본 연구에서는 I2O의 46개 과제 외에 다른 부처의 SW 관련 과제 5개를 포함하여 총 51개 과제를 DARPA의 후보 Pool로 선정하고 검토했다. 대표적인 DARPA의 SW 관련 과제들은 다음과 같다.

11) <부록 7> 참조

12) IARPA 사업으로는 ALADDIN, JANUS 등이 있음

13) IARPA 사업으로는 BABEL 등이 있음

〈표 11〉 대표적인 DARPA의 SW관련 과제 목록

사업	Data	Analytics	Language	Automation	Autonomy	AI
AIDA ¹⁴⁾	○	○	○		○	○
ADAMS ¹⁵⁾		○				○
Assured Autonomy					○	
Big Mechanism	○			○		
BOLT ¹⁶⁾	○	○	○		○	○
BRASS ¹⁷⁾				○		
Causal Exploration ¹⁸⁾		○				
CwC ¹⁹⁾	○				○	○
D3M ²⁰⁾	○	○		○		
DEFT ²¹⁾	○	○	○		○	○
Deep Purple ²²⁾	○					○
EdgeCT ²³⁾		○				
XAI ^{24)주}	○			○	○	○
FLA ²⁵⁾					○	
Insight		○		○		
LORELEI ²⁶⁾	○	○	○	○	○	○
MUSE ^{27)주)}	○	○				○
PPAML ^{28)주}	○					○
RATS ²⁹⁾			○	○		
SD2 ³⁰⁾	○	○				
Transparent Computing				○		
VET ³¹⁾				○		
World Modelers	○			○		○
XDATA	○	○				

주) Programming Language 관련 **과제**

-
- 14) Active Interpretation of Disparate Alternatives
 - 15) Anomaly Detection at Multiple Scales
 - 16) Broad Operational Language Translation
 - 17) Building Resource Adaptive Software Systems
 - 18) Causal Exploration of Complex Operational Environments
 - 19) Communicating with Computers
 - 20) Data-Driven Discovery of Models
 - 21) Deep Exploration and Filtering of Text
 - 22) Deep Purposeful Learning
 - 23) Edge-Directed Cyber Technologies for Reliable Mission Communication
 - 24) Explainable Artificial Intelligence
 - 25) Fast Lightweight Autonomy
 - 26) Low Resource Languages for Emergent Incidents
 - 27) Mining and Understanding Software Enclaves
 - 28) Probabilistic Programming for Advancing Machine Learning
 - 29) Robust Automatic Transcription of Speech
 - 30) Synergistic Discovery and Design
 - 31) Vetting Commodity IT Software and Firmware

2. IARPA

IARPA는 2006년 새롭게 설립된 기관으로 CIA, FBI 등으로 구성된 정보 분야 기관들의 파괴적 혁신의 필요를 충족시키기 위해 탄생했다. IARPA는 정보 분야로 그 영역으로 한정하고 있다. 핵심 관심 영역은 데이터 생명주기와 관련된 4가지로 분석 (analysis), 예측력(anticipatory intelligence), 수집(collection), 컴퓨팅(computing)이다. 총 37개 과제가 진행 중에 있는데, 전체 목록은 아래의 표와 같다.³²⁾

<표 12> 현재 진행 중인 IAPRA 과제

영역	과제
Collection	△FELIX(Finding Engineering-Linked Indicators) △Ithildin △Proteos △Amon-Hen △UnderWatch △Odin (Thor/Loki) △MOSAIC(Multimodal Objective Sensing to Assess Individuals with Context) △Fun GCAT(Functional Genomic and Computational Assessment of Threats) △MAEGLIN(Molecular Analyzer for Efficient Gas-phase Low-power Interrogation) △SILMARILS(Standoff ILluminator for Measuring Absorbance and Reflectance Infrared Light Signatures) △HFGeo(High Frequency Geolocation) △SHARP(Strengthening Human Adaptive Reasoning and Problem-Solving) △SLICE(Signal Location in Complex Environments)
Computing	△SuperCalbes △VirtUE(Virtuous User Environment) △SuperTools △HECTOR(Homomorphic Encryption Computing Techniques with Overhead Reduction) △RAVEN(Rapid Analysis of Various Emerging Nanoelectronics) △LogiQ(Logical Qubits) △QEO(Quantum Enhanced Optimization) △MicrONS(Machine Intelligence from Cortical Networks) △C3(Cryogenic Computing Complexity) △TIC(Trusted Integrated Chips)
Analysis	△DIVA(Deep Intermodal Video Analytics) △CORE 3D △Janus △Finder △MATERIAL(Machine Translation for English Retrieval of Information in Any Language) △Babel △Aladdin Video
Anticipatory Intelligence	△FOCUS(Forecasting Counterfactuals in Uncontrolled Settings) △HFC(Hybrid Forecasting Competition) △Mercury △CREATE(Crowdsourcing Evidence, Argumentation, Thinking and Evaluation) △SCITE(Scientific advances to Continuous Insider Threat Evaluation) △CAUSE(Cyber-attack Automated Unconventional Sensor Environment) △FUSE(Foresight and Understanding from Scientific Exposition)

32) 자세한 사업 내용은 <부록 8> 참조

3. NSF

미국 NSF는 한 해 70억 달러가 넘게 R&D 투자를 하는 중요한 기관이다. 2016 회계연도의 예산은 71억 2,400만 달러였다. 수주기관과 지원방식별 구분은 다음 표와 같다.

<표 13> NSF의 수주기관 및 지원방식별 지원 금액 및 비중 (2016 회계연도 기준)

구 분		금액 (백만달러)	비중 (%)
수주기관	학계	5,420	76
	산업계	1,068	15
	FFRDC ³³⁾	223	3
	기타 ³⁴⁾	412	6
	소계	7,124	100
지원방식	Grants	5,245	74
	CA ³⁵⁾	1,541	22
	Contracts	337	5
	소계	7,124	100

NSF는 컴퓨터 과학을 담당하는 최상위 부서인 CISE(Directorate for Computer & Information Science & Engineering)를 두고 있다. CISE는 3개의 디비전(Division)과 1개의 오피스(Office)를 산하에 두고 있으며, 이들 하위 부서들은 각자 담당하는 영역에서 사업을 운영하고 있다. 또한 CISE의 하위 부서를 아우르는 범CISE사업도 있고, 다른 최상위 부서(Directorate)와 협력하는 사업도 있다. 본 연구에서 검토한 3개 디비전과 1개 오피스에 담당하는 주요 사업 목록 다음과 같다.³⁶⁾

33) Federally Funded Research and Development Centers

34) 연방정부, 주정부, 시정부, 비영리기관, 해외기관 등

35) Cooperative Agreements

36) CISE 부서의 전체 사업 목록은 <부록9> 참조

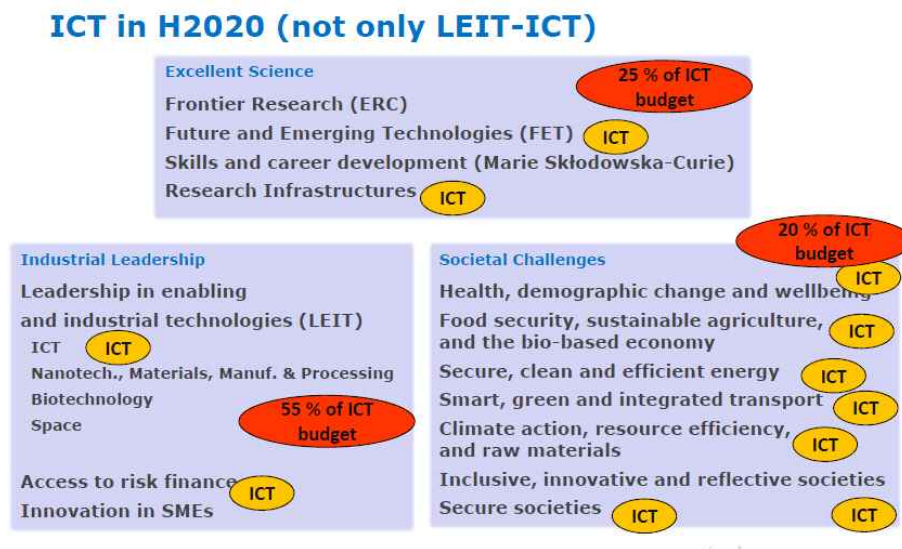
〈표 14〉 NSF CISE부서의 주요 사업

Division of Information & Intelligence Systems (IIS)의 핵심 사업 <ul style="list-style-type: none"> · Cyber-Human Systems (CHS) · Information Integration and Informatics (III) · Robust Intelligence (RI)
Division of Computer and Network Systems (CNS)의 핵심 사업 <ul style="list-style-type: none"> · Computer Systems Research (CSR) · Networking Technology and Systems (NeTS)
Division of Computing and Communication Foundations (CCF)의 핵심 사업 <ul style="list-style-type: none"> · Algorithmic Foundations (AF) · Communications and Information Foundations (CIF) · Software and Hardware Foundations (SHF)
CISE 통합 연구 <ul style="list-style-type: none"> · Information Technology Research · Special Projects
다른 최상위 부서와 협력하는 융합 연구 <ul style="list-style-type: none"> · Critical Techniques, Technologies and Methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering (BIGDATA) · Cyber-Physical Systems (CPS) · Major Research Instrumentation (MRI) · National Robotics Initiative (NRI) · Smart and Autonomous Systems (S&AS) *2016년 공지된 신규 사업 · Smart and Connected Communities (S&CC) · Smart and Connected Health (SCH) · Cyberlearning and Future Learning Technologies (Cyberlearning) · Secure and Trustworthy Cyberspace (S&TC)
CISE의 변혁적 연구(transformative research) <ul style="list-style-type: none"> · Experimental Expedition · Expeditions in Computing · Early-concept Grants for Exploratory Research (EAGER)

4. EU

EU의 R&D 프로그램인 H2020의 경우는 세 가지 목표를 기준으로 구성되어 있다. 과학적 탁월함(excellent science)과 산업적 리더십(industrial leadership), 사회문제 해결(social challenges)을 최상위 목표로 두고 있다. 각 목표에는 하위 목표를 두고 있는데, 전체 구조는 아래의 그림과 같다. 본 연구에서는 ICT와 관련된 사업 중 CPS, 빅데이터 분석, HPC, 클라우드 컴퓨팅, OSS 관련 과제를 중심으로 검토하였다.

<그림 6> EU H2020 내 ICT 관련 사업



출처: Khalil Rouhana (2016)³⁷⁾

37) Khalil Rouhana (2016) ICT in H2020 In support of the EU digital policies

제2절 예비 후보 선발

미국의 DARPA와 IARPA, NSF와 유럽의 H2020에서 현재 진행하고 있는 과제들로부터 본 연구의 목적인 중장기 대형 소프트웨어 R&D과제 발굴을 위한 예비 후보들을 선발했다. 이때 기술 선도국과의 기술 격차 해소 차원에서 개별 과제가 의미가 있는지를 판단해야하기 때문에 다음 세 가지 질문을 선발 기준으로 사용했다.

- 해당 과제가 성공했을 때 우리나라와의 기술 격차가 얼마나 더 확대될 것인가?
- 해당 과제의 목표 또는 목적이 달성될 가능성이 높은가?
- 기술 격차가 확대되는 것을 막기 위해 우리나라의 입장에서 해당 과제에 대응해야 하는가? (단, 대응이 동일한 과제를 기획해서 수행한다는 것을 의미하지는 않음)

총 3,332개의 대상 과제 중에서 1,112개를 검토하였으며 그 중 109개를 예비 후보로 선발했다. 전체 대상 대비 선발 비율은 3.24%이다. 이때 검토는 최근 시작한 과제와 규모가 큰 과제 중심으로 이루어졌다.

〈표 15〉 예비 후보 선발 대상 및 선발 비율

기관		전체 대상(개)	검토 진행(개)	예비 후보(개)	선발 비율(%)
DARPA		219	51	9	4.11
IARPA		37	21	4	10.81
NSF	CPS	350	90	6	1.71
	빅데이터	257	90	10	3.89
	HPC	312	90	10	3.21
	클라우드	730	180	11	1.51
	OSS	287	90	14	4.88
H2020	CPS	114	100	10	8.77
	빅데이터	399	100	8	2.01
	HPC	215	100	10	4.65
	클라우드	244	100	8	3.28
	OSS	168	100	8	4.76
합계		3,332	1,112	108	3.24

1. DARPA 사업 기준 예비 후보

DARPA에서 진행하고 있는 219개 과제 중에서 소프트웨어와 밀접하게 관련된 59개 과제를 대상으로 검토하였다. 검토에는 앞에서 밝힌 세 가지 기준을 적용하여 은 총 9개 과제를 예비 후보로 선발하였다. 선발된 과제의 수행기간과 예산은 다음의 표와 같다.³⁸⁾ 기간은 48개월 내외이며, 연도별 예산은 대부분 1,000만 달러가 넘는다.

<표 16> DARPA 예비 후보의 과제 수행 기간과 예산

사업	기간			예산 (만달러)				
	시작	종료	개월	2014	2015	2016	2017	2018
AIDA ³⁹⁾	'18.	-	54	-	-	-	550	1,730
XAI ⁴⁰⁾	'17.5	'21.4	48	-	-	-	1,100	2,384
D3M ⁴¹⁾	'17.5	'21.4	48	-	-	-	2,025	2,684
SD2 ⁴²⁾	'17.8	-	-	-	-	-	1,300	2,100
PPAML ⁴³⁾	'13.10	'17.7	46	1,022	1,361	1,119	931	-
MUSE ⁴⁴⁾	'14.8	'18.11	52	450	800	1,207	1,300	1,300
SAREWARE ⁴⁵⁾	'15.8	'19.7	48	-	1,000	1,130	1,032	996
BRASS ⁴⁶⁾	'15.10	'19.9	48	-	-	1,734	1,742	1,745
HIVE ⁴⁷⁾	'17.6	'21.9	54	-	-	400	1,671	1,992

38) 각 사업별 상세 내용은 <부록 10> 참조

39) 멀티모달(multi-modal) 학습 및 통합 추론

40) Explainable Artificial Intelligence (XAI) 이해 가능한 딥러닝

41) Data-Driven Discovery of Models (D3M) 도메인 전문가 대상의 데이터 자동 분석 도구

42) Synergistic Discovery and Design (SD2) 데이터 기반 과학 지식 및 모델 발견 도구

43) Probabilistic Programming for Advancing Machine Learning (PPAML) 기계 학습 전용 확률 프로그래밍 언어 개발

44) Mining and Understanding Software Enclaves (MUSE) 대용량의 오픈 소스 소프트웨어 코드 학습을 통한 소프트웨어 코드 자동 분석 및 오류 제거

45) 소프트웨어 코드 역공학 방지 솔루션

46) Building Resource Adaptive Software Systems (BRASS) 소프트웨어 운영 환경 상의 논리적 또는 물리적 자원 변화에 적응하여 장기간 동작할 수 있는 소프트웨어 (기초연구)

47) Hierarchical Identify Verify Exploit (HIVE) 대용량 그래프 데이터 분석 전용 프로세서 (폰노이만 구조 탈피)

2. IARPA 예비 후보

IARPA에서 진행하고 있는 37개 과제 중에서 소프트웨어와 밀접하게 관련된 21개 과제를 대상으로 검토하여 총 4개 과제를 예비 후보로 선발했다.⁴⁸⁾ 선발된 과제의 수행 기간은 DARPA 과제들과 마찬가지로 48개월 내외인데, 대체로 3단계로 구분해서 진행한다. MATERIAL과제에서는 2단계와 3단계의 경우, 경쟁을 통해 선별적으로 과제 참여 여부가 결정된다.

<표 17> IARPA 예비 후보의 과제 공고와 수행 기간

사업	사업공고		기간 (개월)	비고
	공고 번호	일자		
MATERIAL ⁴⁹⁾	16-11	2017.1.19	46	18(기본) + 16(선택) + 12(선택)
DIVA ⁵⁰⁾	16-13	2016.9.13	48	18 + 18 + 12
Odin (Thor/Loki) ⁵¹⁾	16-04	2016.6.15	48	18 + 18 + 12
SuperTools ⁵²⁾	16-03	2016.6.07	60	24 + 24 + 12

3. NSF 예비 후보

NSF에서 진행하고 있는 과제의 수가 많기 때문에 분야별로 구분해서 예비 후보를 선발하였다. 분야는 사이버물리시스템(CPS)과 빅데이터 분석, 고성능컴퓨팅(HPC), 클라우드 컴퓨팅, 오픈 소스 소프트웨어(OSS)이다.

① 사이버물리시스템(CPS)

NSF에서 진행하고 있는 과제들 중 CPS와 관련된 350개 과제가 확인되었다. 그 중에서 최근에 시작됐고 100만 달러 이상의 대형 연구를 중심으로 90개 과제를 검토하여 6개 예비 후보를 선발했다. 검토에는 앞에서 밝힌 세 가지 기준이 이용됐다.

48) 사업별 상세내용은 <부록 11> 참조

49) Machine Translation for English Retrieval of Information in Any Language (MATERIAL) 영어 질문에 대해 다른 언어로 된 자료를 검색하여 요약한 후에 영어로 번역하여 결과 제공

50) Deep Intermodal Video Analytics (DIVA) 복수의 실시간 촬영 카메라 환경에서의 객체와 동작 인식

51) Odin 알려진 또는 새로운 생체정보 도용 수법 탐지 (2개의 세부사업인 Thor와 Loki로 구분하여 추진)

52) SuperTools 초전도 전자기기용 전자 설계 자동화(EDA: Electronic Design Automation) 도구

[예비후보1] Building a Smart City Economy and Information Ecosystem to Motivate Pro-Social Transportation Behavior⁵³⁾

[예비후보2] Security and Privacy-Aware Cyber-Physical Systems⁵⁴⁾

[예비후보3] Secure Perception for Autonomous Systems⁵⁵⁾

[예비후보4] A Cyber-Physical System for Monitoring, Analysis and Mitigation of Urban Noise Pollution⁵⁶⁾

[예비후보5] Foundations of Resilient CybEr-Physical Systems (FORCES)⁵⁷⁾

[예비후보6] Correct-by-Design Control Software Synthesis for Highly Dynamic Systems⁵⁸⁾

선발된 과제의 수행기간과 예산, 수행의 주관기관은 다음의 표와 같다. 기간은 36개월과 60개월로 구분되며, 각 과제의 예산은 대부분 100만 달러 이상인데, 특히 예비후보5와 예비후보6은 복수의 연구기관에서 수행하며 예산 규모도 각각 900만 달러 이상과 260만 달러이다.

<표 18> NSF CPS 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산 (천달러)	주관기관
	시작	종료 ^{주)}	개월		
예비후보1	'17.9	'20.8	36	1,440	Univ. of Pittsburgh
예비후보2	'17.9	'20.8	36	1,125	Univ. of Pennsylvania
예비후보3	'15.4	'20.3	60	500	Univ. of Texas at Austin
예비후보4	'16.8	'21.7	60	2,015	New York Univ.
예비후보5	'13.4	'18.3	60	2,150	MIT
				3,600	Univ. of California-Berkeley
				1,945	Vanderbilt Univ.
				1,300	Univ. of Michigan Ann Arbor
예비후보6	'13.4	'18.3	60	1,000	UCLA
				1,600	Univ. of Michigan Ann Arbor

주) NSF에서는 종료기간을 고정하지 않고 예산 정보를 제시함

53) 도시 대중교통 운행 및 좌석 정보를 이용자들에게 실시간 전달하고 그 정보를 이용할 수 있는 장터 (캘리포니아주 오클랜드시와 펜실바니아주 피츠버그시 도심에 실험적 적용)

54) 악성 공격으로부터 안전하고 차등화된 프라이버시 관리가 가능한 CPS

55) 드론, 자율주행차 등의 자율시스템의 보안 공격을 인식할 수 있는 방법

56) 도시의 소음 공해를 모니터링하고 분석하여 해소시킬 수 있는 CPS

57) 복원력이 있는 CPS 제어 알고리즘 (기초연구)

58) 실시간 대응이 필요한 환경에서 작동하는 시스템 내 복수의 제어 소프트웨어의 통합

② 빅데이터

NSF에서 진행하고 있는 과제들 중 빅데이터와 관련된 257개 과제가 확인되었다. 그 중에서 최근에 시작한 것들, 그리고 100만 달러 이상의 대형 연구를 중심으로 90개 과제를 검토하여 다음과 같은 10개 예비 후보를 선발하였다. 검토는 앞에서 밝힌 세 가지 기준을 의거해서 이루어졌다.

[예비후보01] A Formal Foundation for Big Data Management⁵⁹⁾

[예비후보02] Discovery and Social Analytics for Large-Scale Scientific Literature⁶⁰⁾

[예비후보03] User Driven Architecture for Data Discovery⁶¹⁾

[예비후보04] Human and Machine Co-Processing⁶²⁾

[예비후보05] Big Data Modeling and Analysis with Depth and Scale⁶³⁾

[예비후보06] Addressing the two V's of Veracity and Variety in Big Data⁶⁴⁾

[예비후보07] Large-Scale Transductive Learning from Heterogeneous Data Sources⁶⁵⁾

[예비후보08] Study of a Cyber-Enabled Social Computing Framework for Improving Practice in Online Computing Communities⁶⁶⁾

[예비후보09] Hype Cycles of Scientific Innovation⁶⁷⁾

[예비후보10] Open Source Supply Chains and Avoidance of Risk (OSCAR): An Evidence Based Approach to Improve FLOSS Supply Chains⁶⁸⁾

59) 빅데이터 탐색에 소요되는 시간을 줄일 수 있는 미들웨어 개발 (단, 기존 솔루션과 연동 가능)

60) 개인의 과학 분야 리퍼지토리 사용 패턴을 이용한 데이터 탐색 속도 향상

61) 사용자 중심의 데이터 검색 아키텍처: 최소한의 노력으로 여러 과학자 커뮤니티에서 관리하고는 데이터의 글로벌 인덱스 생성

62) 도메인 전문가의 대용량 데이터 탐색 보조: 기계학습을 통해 최소한의 인터랙션을 통해 도메인 전문가의 요구사항을 이해하여 도메인 전문가가 다룰 수 있는 크기의 데이터 서브셋 생성

63) 대용량 복잡 데이터 셋을 빠르게 이해하는데 필요한 통계-논리 결합 모델을 생성하고 평가하고 수정할 수 있는 프로그래밍 언어 및 시스템

64) 데이터의 일관성, 신뢰성, 출처 등의 데이터 품질을 평가하고 데이터 획득 단계에서부터 곧바로 데이터의 책임 소재와 품질 관리 주체를 설정할 수 있는 데이터 평가 엔진(DVE: Data Valuation Engine)

65) 그래프 모델을 이용한 대용량 이종 데이터 학습

66) GitHub와 같은 온라인 협업 커뮤니티와 MOOC와 온라인 학습 커뮤니티 데이터를 결합하여 참여자들의 행동 패턴에 기초한 복수의 그룹을 구분한 후 그룹별 요구사항이 충족되는 방식을 확인

67) 1990년부터 2016년까지의 과학논문, 정부지원과제, 특허 등의 데이터를 통합하여 미국의 과학발전 패턴 분석 (주제 모델링, 자연어처리 등 활용)

68) 오픈소스 코드와 기여 활동에 관한 빅데이터를 분석하여 오픈소스소프트웨어의 공급사슬을 분석하여 공급

선발된 과제의 수행기간과 예산, 수행의 주관기관은 다음의 표와 같다. 이들 과제들의 시작 시기는 2013년부터 2016년까지 다양하며, 종료기간 또한 2017년부터 2020년까지 다양하다. 기간은 36개월과 48개월, 60개월로 다양하며, 각 과제의 예산은 모두 100만 달러가 넘는다.

〈표 19〉 NSF 빅데이터 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산 (천달러)	주관기관
	시작	종료 ^{주)}	개월		
예비후보1	'13.1	'17.12	60	2,967	Univ. of Washington
예비후보2	'13.1	'17.12	60	1,005	Rutgers Univ. New Brunswick
예비후보3	'14.9	'18.04	44	1,485	Corporation for NRI ⁶⁹⁾
예비후보4	'14.9	'18.8	48	1,397	Univ. of Wisconsin-Madison
예비후보5	'14.8	'18.7	48	1,500	SUNY at Stony Brook
예비후보6	'14.9	'18.8	48	1,000	Univ. of Notre Dame
예비후보7	'16.1	'19.12	48	1,205	Carnegie-Mellon Univ.
예비후보8	'16.1	'19.12	48	1,599	Carnegie-Mellon Univ.
예비후보9	'16.9	'19.8	36	1,253	Stanford Univ.
예비후보10	'16.9	'20.8	48	1,300	Univ. of Tennessee Knoxville
				429	Carnegie-Mellon Univ.

주) NSF에서는 종료기간을 고정하지 않고 예상 정보를 제시함

③ HPC

NSF에서 진행하고 있는 과제들 중 HPC와 관련된 312개 과제가 확인되었다. 앞과 동일한 조건으로 90개 과제를 검토하여 다음과 같은 10개 예비 후보를 선발했다. 검토는 앞에서 밝힌 세 가지 기준을 이용해서 이루어졌다. 선발된 과제의 수행기간과 예산, 주관기관은 다음의 표와 같다. 과제들의 시작 시기는 2014년부터 2017년까지 다양하며, 종료기간 또한 2018년부터 2021년까지 다양하다. 기간은 36개월과 48개월, 60개월

사슬상의 위험을 최소화시킬 수 있는 방안 마련

69) National Research Initiatives

로 다양하며, 각 과제의 예산은 50만 달러에서 500만 달러 이상까지 다양하다.

[예비후보01] Compute on Data Path: Combating Data Movement in High Performance Computing⁷⁰⁾

[예비후보02] Middleware and High Performance Analytics Libraries for Scalable Data Science⁷¹⁾

[예비후보03] Performance Application Programming Interface for Extreme-Scale Environments (PAPI-EX)⁷²⁾

[예비후보04] Expanding Volunteer Computing⁷³⁾

[예비후보05] Building High-Availability Data Capabilities in Data-Centric Cyberinfrastructure⁷⁴⁾

[예비후보06] Hierarchical Tuning of Floating-Point Computations⁷⁵⁾

[예비후보07] Enhancing the Open MPI Software for Next Generation Architectures and Applications⁷⁶⁾

[예비후보08] Ubiquitous Access to Transient Data and Preliminary Results via the SeedMe Platform⁷⁷⁾

[예비후보09] Sustainable Open-Source Quantum Dynamics and Spectroscopy Software⁷⁸⁾

[예비후보10] High Performance and Scalable Fabric Analysis, Monitoring and Introspection Infrastructure for HPC and Big Data⁷⁹⁾

70) 데이터와 컴퓨팅의 결합을 객체로 인식한 데이터 이동 경로 기반의 컴퓨팅 모델 (가능성 검증)

71) HPC 환경에서의 대용량 데이터 분석을 위한 소프트웨어 미들웨어와 라이브러리 설계 및 구축

72) 시장의 지배적인 마이크로프로세서의 하드웨어 성능 분석 인터페이스인 PAPI의 차세대 버전 개발 (인터코어 리소스 등 최신 컴퓨팅 아키텍처의 변화 반영)

73) PC, 스마트폰 등의 컴퓨팅 자원을 기부 받는 자원제공자 컴퓨팅(VC: Volunteer Computing)을 기존 HPC 설비에 추가하는 것을 목표로 함 (대표적인 소프트웨어 프레임워크인 BOINC 이용)

74) 현행 HPC 파일 시스템과 하둡 HDFS의 중앙 집중화된 인덱싱 한계를 극복하고 엑사스케일 파일 시스템으로 확장 (교육 훈련과 협력 연구 포함)

75) 부동소수점 계산에서의 비효율을 줄여 HPC부터 모바일기기 모든 컴퓨팅기기의 계산 성능 향상

76) Open MPI 소프트웨어 라이브러리 개선: 새로운 기능 지원과 새로운 아키텍처 지원

77) 웹 기반의 연구 데이터 공유 플랫폼인 SeedMe를 이용하여 일시 저장 데이터와 예비 연구 결과의 전송과 공유

78) 계산 화학 소프트웨어 패키지인 크로누스 쿼텀(Chronus Quantum)에 양자 역학 시뮬레이션 기능 추가

79) HPC의 빅데이터 분석 성능을 향상시키기 위한 종합 소프트웨어 인프라에 탑재될 미들웨어와 툴

〈표 20〉 NSF HPC 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산 (천달러)	주관기관
	시작	종료	개월		
예비후보1	'14.6	'18.5	48	1,016	Texas Tech Univ.
예비후보2	'14.10	'19.9	60	5,154	Indiana Univ.
예비후보3	'15.9	'19.8	48	2,126	Univ. of Tennessee Knoxville
예비후보4	'17.5	'20.4	36	500	Univ. of Texas at Austin
예비후보5	'17.8	'20.7	36	500	Univ. of Virginia Main Campus
예비후보6	'17.8	'20.7	36	1,200	Univ. of Utah
예비후보7	'17.6	'21.5	48	1,566	Univ. of Tennessee Knoxville
예비후보8	'14.10	'18.9	48	1,329	Univ. of California-San Diego
예비후보9	'17.9	'21.8	48	1,500	Univ. of Washington
예비후보10	'17.7	'20.6	36	800	Ohio State Univ.

④ 클라우드

NSF에서 진행하고 있는 과제들 중 클라우드와 관련된 730개 과제가 확인되었다. 그 중 180개 과제를 검토하여 다음과 같은 11개 예비 후보를 선발하였다. 과제들의 기간은 36개월과 48개월, 56개월, 60개월로 다양하며, 각 과제의 예산은 50만 달러에서 1,200만 달러 이상까지 다양하다. 이때 [예비후보7] Jetstream은 연구개발이자 조달구축사업이기도 하다.

[예비후보01] Community Infrastructure To Expand the Frontiers of Cloud Computing Research⁸⁰⁾

[예비후보02] Chameleon: A Large-Scale, Reconfigurable Experimental Environment for Cloud Research⁸¹⁾

[예비후보03] Self-Monitoring Virtual Machines for Performance Guarantees in Public Clouds⁸²⁾

[예비후보04] New Abstractions for First-hop Networking in Cloud Data Centers⁸³⁾

80) 컴퓨터 과학 분야 연구를 위한 실험용 컴퓨팅 클라우드인 Cloudlab의 고도화

81) 클라우드 컴퓨팅 연구용 테스트베드의 고도화

82) 클라우드 서비스 사업자의 도움 없이 서비스 이용자들이 이용 자원의 제공 상태를 자가 모니터링

83) 다중 애플리케이션 서버 실행 환경에서 고속 통신을 위한 서버 운영체제와 네트워크 인터페이스 카드(NIC)

- [예비후보05] Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation⁸⁴⁾
- [예비후보06] An Architecture for Scalable Verifiable Computing⁸⁵⁾
- [예비후보07] Language-Hardware Co-Design for Practical and Verifiable Information Flow Control⁸⁶⁾
- [예비후보08] Jetstream: A Self-Provisioned, Scalable Science and Engineering Cloud Environment⁸⁷⁾
- [예비후보09] Enabling Flexible Middlebox Processing in the Cloud⁸⁸⁾
- [예비후보10] A Computing Cloud for Graphical Simulation⁸⁹⁾
- [예비후보11] NSF Net-centric and Cloud Software and Systems⁹⁰⁾

의 새로운 알고리즘과 프로그래밍 추상화

- 84) 뉴욕주와 캘리포니아주 대학 캠퍼스들 대상의 페더레이티드 클라우드 컴퓨팅용 데이터 분석 및 관리 소프트웨어 (코넬대학 주도)
- 85) 클라우드 컴퓨팅 서비스를 적절하게 제공하고 있음을 증명할 수 있는 프로토콜
- 86) 모든 데이터 흐름을 추적할 수 있는 컴퓨팅 시스템 설계 방법론
- 87) 클라우드 기반의 슈퍼컴퓨팅 시스템
- 88) 기업의 클라우드로의 전환할 때 기존 성능과 보안을 충족하면서 동시에 비용을 절감할 수 있는 로드 밸런서, 침입 탐지, WAN 최적화 툴 등 미들박스(middleboxes)
- 89) 클라우드 기반의 그래픽 시뮬레이션 소프트웨어인 Nimbus를 오픈 소스 기반의 대규모 복잡한 그래픽을 시뮬레이션할 수 있는 PhysBAM 사용자들이 이용할 수 있게 고도화
- 90) NSF가 지원하는 산학협력연구센터 중 하나인 Net-centric and Cloud Software and Systems Industry/University Cooperative Research Center (NCSS I/UCRC)의 2단계 사업 (1단계 사업: 2009~2014.2)

〈표 21〉 NSF 클라우드 컴퓨팅 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산 (천달러)	주관기관
	시작	종료	개월		
예비후보1	'17.10	'20.9	36	5,834	Univ. of Utah
예비후보2	'17.10	'20.9	36	3,102	Univ. of Chicago
				2,368	Univ. of Texas at Austin
				100	Northwestern Univ.
예비후보3	'17.10	'20.9	36	500	Ohio State Univ.
예비후보4	'17.9	'20.8	36	500	Univ. of Wisconsin-Madison
예비후보5	'15.10	'20.9	60	6,115	Cornell Univ.
예비후보6	'15.7	'18.6	36	1,163	Univ. of Maryland College Park
예비후보7	'15.7	'19.6	48	1,200	Cornell Univ.
예비후보8	'14.12	'19.1	60	12,341	Indiana Univ.
예비후보9	'14.1	'18.8	56	500	Carnegie-Mellon Univ.
예비후보10	'14.8	'18.7	48	1,093	Stanford Univ.
예비후보11	'14.4	'19.3	60	1,093	Univ. of North Texas

⑤ OSS(Open Source Software)

NSF에서 진행하고 있는 과제들 중 OSS와 관련된 287개 과제가 확인되었다. 그 중 90개 과제를 검토하여 14개 예비 후보를 선발하였다. 기간은 36개월이 대부분인데, 48개월과 60개월도 있다, 각 과제의 예산은 40만 달러에서 650만 달러 이상까지 다양하다.

[예비후보01] Chameleon: A Large-Scale, Reconfigurable Experimental Environment for Cloud Research⁹¹⁾⁹²⁾

[예비후보02] Gunrock: High-Performance GPU Graph Analytics⁹³⁾

[예비후보03] Entangled Quantum Dynamics in Closed and Open Systems, an Open Source Software Package for Quantum Simulator Development and Exploration of Synthetic Quantum Matter⁹⁴⁾

[예비후보04] An Ecosystem of Reusable Image Analytics Pipelines⁹⁵⁾

[예비후보05] Enhancing Context-Awareness and Personalization for

91) 클라우드 컴퓨팅 연구용 테스트베드의 고도화

92) NSF 클라우드 컴퓨팅 과제 기준 [예비후보02]와 동일

93) GPU를 이용하는 그래픽 분석 라이브러리

94) 양자 시뮬레이터 개발용 오픈소스소프트웨어 패키지 (합성 양자 물질 분야 시범 적용)

95) 대용량 이미지 데이터 분석용 오픈소스 프레임워크 (빅데이터 관리·분석 시스템인 Myria와 통합)

Intensively Adaptive Smoking Cessation Messaging
Interventions⁹⁶⁾

- [예비후보06] Flexible Resource Management and Coordination Schemes for Lightweight, Rapidly Deployable OS/Rs⁹⁷⁾
- [예비후보07] Creating Natural Data Visualization and Analysis Environments⁹⁸⁾
- [예비후보08] Enabling Sensor-Rich Vehicular Applications with Edge Computing⁹⁹⁾
- [예비후보09] The SimCardio Open Source Multi-Physics Cardiac Modeling Package¹⁰⁰⁾
- [예비후보10] BONSAI: An Open Software Infrastructure for Parallel Autotuning of Computational Kernels¹⁰¹⁾
- [예비후보11] Software Framework for Research in Semi-Autonomous Teleoperation¹⁰²⁾
- [예비후보12] Robust Algorithms for an Open Source Software Reliability Tool¹⁰³⁾
- [예비후보13] Enabling Technologies for 21st Century Entity Matching Applications¹⁰⁴⁾
- [예비후보14] Open Gateway Computing Environments Science Gateways Platform as a Service (OGCE SciGaP)¹⁰⁵⁾

96) 메시징 기반의 금연 지원 오픈소스소프트웨어 시스템

97) 특정 애플리케이션에 최적화되고 빠르게 설정할 수 있는 경량 운영체제와 실행환경

98) 데이터 시각화 및 시각화 기반 분석을 위한 멀티모달 인터랙션 기반의 사용자 인터페이스(UI)

99) 다수 센서로부터 대용량 데이터를 생성하는 자동차용 엣지 컴퓨팅

100) 다차원 심장 모델링을 위한 오픈소스소프트웨어 패키지

101) 이 기종 프로세서 기반 HPC 시스템에서의 병렬 자동 튜닝에 의한 소프트웨어 최적화

102) 원격실험실 기구의 반자율동작을 지원하는 소프트웨어 프레임워크

103) 수치적으로 안정적인 오픈소스 기반의 신뢰성 툴

104) 빅데이터 분석을 위한 엔터티 매칭 애플리케이션 (예: UW-Madison = Univ of Wisc madison)

105) 연구자 중심의 가상 연구 환경인 사이언스 게이트웨이(Science Gateway)의 확장

〈표 22〉 NSF OSS 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산 (천달러)	주관기관
	시작	종료	개월		
예비후보1	'17.10	'20.9	36	3,102	Univ. of Chicago
				2,368	Univ. of Texas at Austin
				100	Northwestern Univ.
예비후보2	'17.10	'20.9	36	400	Univ. of California-Davis
예비후보3	'17.9	'20.8	36	500	Colorado School of Mines
예비후보4	'17.9	'20.8	36	500	Univ. of Washington
예비후보5	'17.10	'20.9	36	640	Univ. of Massachusetts Amherst
				196	Univ. of Memphis
				163	Univ. of Massachusetts Medical School
예비후보6	'17.8	'20.7	36	250	Illinois Institute of Technology
				250	Univ. of Pittsburgh
예비후보7	'17.11	'20.10	36	494	Georgia Tech Research Corporation
예비후보8	'17.10	'20.9	36	499	Univ. of Michigan Ann Arbor
예비후보9	'17.9	'21.8	48	1,431	Stanford Univ.
				944	Univ. of California-Berkeley
예비후보10	'16.11	'19.10	36	500	Univ. of Tennessee Knoxville
예비후보11	'16.10	'19.9	36	978	Johns Hopkins Univ.
				550	Worcester Polytechnic Institute
				480	Univ. of Washington
예비후보12	'15.10	'18.9	36	150	Univ. of Massachusetts, Dartmouth
예비후보13	'16.9	'20.8	48	1,086	Univ. of Wisconsin-Madison
예비후보14	'13.10	'18.9	60	2,520	Indiana Univ.
				1,742	Univ. of California-San Diego
				600	UT Health Science Center San Antonio

4. H2020 예비 후보

① CPS

H2020에서 진행하고 있는 과제들 중 CPS와 관련된 114개 과제가 확인되었다. 그 중에서 최근에 시작한 것들, 그리고 100만 달러 이상의 대형 연구를 중심으로 100개 과제를 검토하여 다음과 같은 10개 예비 후보를 선발하였다. 검토는 앞에서 밝힌 세 가지 기준을 이용해서 이루어졌다.

- Dependability Engineering Innovation for CPS (DEIS)¹⁰⁶⁾
- Platform for Open Development of Systems of Artificial Intelligence (Bonseyes)¹⁰⁷⁾
- Unifying Control and Verification of Cyber-Physical Systems (UnCoVerCPS)¹⁰⁸⁾
- SAFety and secURity by design for interconnected mixed-critical cyber-physical systems (SAFURE)¹⁰⁹⁾
- Testing Cyber-Physical Systems under Uncertainty: Systematic, Extensible, and Configurable Model-based and Search-based Testing Methodologies (U-Test)¹¹⁰⁾
- European Network of competencies and platforms for Enabling SME from any sector building Innovative CPS products to sustain demand for European manufacturing (EuroCPS)¹¹¹⁾
- Trusted Apps for open CPS (TAPPS)¹¹²⁾
- Integrated Modelling, Fault Management, Verification and Reliable Design Environment for Cyber-Physical Systems (IMMORTAL)¹¹³⁾
- Agile, eXtensible, fast I/O Module for the cyber-physical era (AXIOM)¹¹⁴⁾
- Towards Cyber-Physical Systems Engineering Tools Interoperability Standardisation (CP-SETIS)¹¹⁵⁾

106) 자동차, 기차, 의료기기 등에서의 부품 또는 단위 시스템의 신뢰성(dependability) 있는 통합

107) 인공지능 시스템 개발을 위한 플랫폼 (데이터 장터, 딥러닝 툴박스, 개발자 참고 플랫폼 등)

108) CPS의 제어 및 오류검증 통합

109) 다양한 수준의 안전이 요구되는 CPS의 안전과 보안 통합 (기존 분리 패러다임 탈피)

110) CPS의 신뢰성(dependability) 향상을 위한 자동화된 모델 기반의 테스트 방법론

111) 유럽연합 중소기업의 CPS 제품개발을 지원하기 위한 플랫폼

112) CPS용 앱을 위한 플랫폼 (앱 스토어 포함)

113) 신뢰성(dependability) 있는 CPS를 위한 오류 관리 및 검증, 모델링, 설계 통합 프레임워크

114) 멀티코어 멀티보드 CPS를 위한 FPGA 기반의 분산 공유 메모리 모듈러 시스템

선발된 과제의 수행기간과 예산, 주관기관은 다음의 표와 같다. 기간은 36개월이 대부분이지만 26개월과 48개월도 있다. 각 과제의 예산은 대부분 300만 유로가 넘지만 예외적으로 70만 유로에 미치는 못하는 경우도 있다. 기업이 주로 주관기관을 담당하고 있다.

〈표 23〉 H2020 CPS 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산(유로)	주관국가	주관기관 유형
	시작	종료	개월			
DEIS	'17.1	'19.12	36	4,889,290	오스트리아	기업
Bonseyes	'16.12	'19.11	36	5,018,025	스위스	기업
UnCoVerCPS	'15.1	'18.12	48	4,932,902	독일	대학
SAFURE	'15.2	'18.1	36	5,231,375	오스트리아	기업
U-Test	'15.1	'17.12	36	3,713,234	노르웨이	기업
EuroCPS	'15.2	'18.1	36	8,186,834	프랑스	연구소
TAPPS	'15.1	'17.12	36	3,885,484	독일	기업
IMMORTAL	'15.3	'18.2	36	3,996,653	에스토니아	대학
AXIOM	'15.2	'18.1	36	3,945,938	이탈리아	대학
CP-SETIS	'15.3	'17.5	26	698,895	독일	기타

② 빅데이터

H2020에서 진행하고 있는 빅데이터와 관련된 399개 과제 중 100개 과제를 검토하여 8개 예비 후보를 선발했다. 선발된 과제의 수행기간과 예산, 주관기관은 다음의 표와 같다. 기간은 36개월이 대부분이지만 26개월과 48개월도 있다. 각 과제의 예산은 대부분 300만 유로가 넘지만 예외적으로 70만 유로에 미치는 못하는 경우도 있다. 기업이 주로 주관기관을 담당하고 있다.

- Perceptive Sentinel: BIG DATA knowledge extraction and re-creation platform (PerceptiveSentinel)¹¹⁶⁾
- API-ecosystem for cross-sectorial exchange of 3D personal data (BodyPass)¹¹⁷⁾
- Data-driven decision making for a more efficient society (BiggerDecisions)¹¹⁸⁾

115) CPS 개발 툴 상호운영성을 위한 표준화

116) 지구 관찰(EO: Earth Observation) 빅데이터 분석 플랫폼

117) 병원에서 수집하고 보관 중인 3차원 개인 데이터의 소비재 시장 이용을 위한 생태계

118) 프라이버시 보호를 요구하는 데이터를 이용하여 의사결정 품질을 높인 셰어마인드(ShareMind)를 글로벌

- MAKing Sustainable development and WELL-being frameworks work for policy analysis (MAKSWELL)¹¹⁹⁾
- A Machine learning enabled Clinical Decision Support System to prevent prescription errors and improve patient safety (MEDAS)¹²⁰⁾
- Scalable Policy-aware linked data arChitecture for prIvacy, trAnsparency and compLIance (SPECIAL)¹²¹⁾
- Scalable Oblivious Data Analytics (SODA)¹²²⁾
- BIG DATA FOR SMART SOCIETY (GATE)¹²³⁾

<표 24> H2020 빅데이터 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산(유로)	주관국가	주관기관 유형
	시작	종료	개월			
PerceptiveSentinel	'18.1	'20.6	30	1,983,187	슬로베니아	기업
BodyPass	'18.1	'20.12	36	2,552,018	스페인	기업
BiggerDecisions	'17.9	'19.2	18	1,281,437	에스토니아	기업
MAKSWELL	'17.11	'20.4	30	934,297	이탈리아	연구소
MEDAS	'17.5	'17.8	4	50,000	이스라엘	기업
SPECIAL	'17.1	'19.12	36	3,991,389	프랑스	기타
SODA	'17.1	'19.12	36	2,980,610	네덜란드	기업
GATE	'17.9	'18.8	12	387,875	불가리아	대학

③ HPC

H2020에서 진행하고 있는 과제들 중 HPC와 관련된 것으로 215개 과제가 확인되었다. 그 중에서 최근에 시작한 것들, 그리고 100만 달러 이상의 대형 연구를 중심으로 100개 과제를 검토하여 다음과 같은 10개 예비 후보를 선발하였다. 선발된 과제의 수행기간은 36개월이 대부분이지만 42개월도 있다. 예산은 500만 유로가 넘는 과제들이 많은데, 이는 다른 분야에 비해서도 많은 편이다. 한편 연구소와 대학이 주로 주관기관 역할을 수행하고 있다.

버전으로 확대

119) GDP의 한계를 넘어서는 사회 발전 지표

120) 기계학습을 활용한 임상 의사결정지원시스템(CDSS)

121) 프라이버시 보호, 투명한 공개, 법규 준수 등이 가능한 링크드 데이터 아키텍처 (대시보드 제공 등)

122) 복수 기관에서 소장하고 있는 프라이버시 민간 데이터를 복수 기관이 협력하여 처리하는 테크닉 (예: 보건의료 분야)

123) 스마트 사회를 위한 빅데이터 연구센터

- Co-designed Innovation and System for Resilient Exascale Computing in Europe: From Applications to Silicon (EuroEXA)¹²⁴⁾
- DEEP-Extreme Scale Technologies (DEEP-EST)¹²⁵⁾
- Transparent heterogeneous hardware Architecture deployment for eNergy Gain in Operation (TANGO)¹²⁶⁾
- European Exascale System Interconnect and Storage (ExaNeSt)¹²⁷⁾
- Next Generation I/O for Exascale (NEXTGenIO)¹²⁸⁾
- Energy-efficient Heterogeneous COmputing at exaSCALE (ECOSCALE)¹²⁹⁾
- MANGO: exploring Manycore Architectures for Next-GeneratiOn HPC systems¹³⁰⁾
- European Exascale Processor Memory Node Design (ExaNoDe)¹³¹⁾
- Exploiting eXascale Technology with Reconfigurable Architectures (EXTRA)¹³²⁾
- Runtime Exploitation of Application Dynamism for Energy-efficient eXascale computing (READEX)¹³³⁾

〈표 25〉 H2020 HPC 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산(유로)	주관국 가	주관기관 유형
	시작	종료	개월			
EuroEXA	'17.9	'21.2	42	19,949,023	그리스	연구소
DEEP-EST	'17.7	'20.6	36	14,998,342	독일	연구소
TANGO	'16.1	'18.12	36	3,199,625	스페인	기업
ExaNeSt	'15.12	'18.11	36	8,442,547	그리스	연구소
NEXTGenIO	'15.10	'18.9	36	8,114,504	영국	대학
ECOSCALE	'15.10	'18.9	36	4,237,398	그리스	연구소
MANGO	'15.10	'18.9	36	5,801,820	스페인	대학
ExaNoDe	'15.10	'18.9	36	8,629,248	프랑스	연구소
EXTRA	'15.9	'18.8	36	3,989,931	벨기에	대학
READEX	'15.9	'18.8	36	3,534,199	독일	대학

124) 400 PFlops급 슈퍼컴퓨팅 시스템 (애플리케이션부터 프로세서까지)

125) 세계 최초의 모듈러 슈퍼컴퓨터 아키텍처

126) 이종 병렬 컴퓨팅에서 전력 소모량 최적화 목적에 자가 적응할 수 있는 소프트웨어

127) 엑사스케일 슈퍼컴퓨터용 인터커넥트와 스토리지

128) 엑사스케일 슈퍼컴퓨터용 I/O (비휘발성 듀얼 인라인 메모리 모듈 이용)

129) 엑사스케일 슈퍼컴퓨터의 에너지 효율 제고를 위한 아키텍처

130) 100 PFlops의 매니코어 슈퍼컴퓨터

131) 엑사스케일 슈퍼컴퓨터용 이종 멀티 칩 모듈(MCM: Multi-Chip-Module)

132) 엑사스케일 슈퍼컴퓨터용 컴퓨팅 노드 재설정을 위한 플랫폼

133) 엑사스케일 슈퍼컴퓨터의 에너지 효율 제고를 위한 애플리케이션 실행 최적화 툴

4.2.4.4. H2020 클라우드 과제 기준 예비 후보

H2020에서 진행하고 있는 클라우드와 관련된 244개 과제 중 100개 과제를 검토하여 다음과 같은 8개 예비 후보를 선발하였다. 선발된 과제의 수행기간은 24개월부터 42개월까지 다양하고, 예산은 모두 100만 유로는 넘지만 500만 유로가 넘는 과제는 없다. 한편 연구소와 대학이 주로 주관기관 역할을 수행하고 있다.

- Novel Enablers for Cloud Slicing (NECOS)¹³⁴⁾
- Adaptive, Trustworthy, Manageable, Orchestrated, Secure, Privacy-assuring, Hybrid Ecosystem for REsilient Cloud Computing (ATMOSPHERE)¹³⁵⁾
- A NOVEL FRAMEWORK FOR MULTI-CLOUD SERVICES DEVELOPMENT, ORCHESTRATION, DEPLOYMENT AND CONTINUOUS MANAGEMENT FOSTERING CLOUD TECHNOLOGIES UPTAKE FROM DIGITAL SMES AND STARTUPS (UNICORN)¹³⁶⁾
- ACTivating resource efficiency and large databases in the CLOUD (ACTiCLOUD)¹³⁷⁾
- Cloud Orchestration at the Level of Application (COLA)¹³⁸⁾
- DITAS: Data-intensive applications Improvement by moving daTA and computation in mixed cloud/fog environment¹³⁹⁾
- Multi-cloud Execution-ware for Large-scale Optimized Data-Intensive Computing (MELODIC)¹⁴⁰⁾
- Cloud Large Scale Video Analysis (Cloud-LSVA)¹⁴¹⁾

134) 경량 슬라이스 정의형 클라우드 (LSDC: Lightweight Slice Defined Cloud)

135) 차세대 신뢰할 수 있는 페더레이티드 대륙간 하이브리드 클라우드 서비스 (예: EU-브라질)

136) 유럽연합 중소기업과 신생기업의 소프트웨어 출시 시간을 줄이고 인지도를 높일 수 있는 멀티 클라우드 서비스

137) 비용 대비 효율적이며 고성능 서비스로서의 데이터베이스(DBaaS) 클라우드 플랫폼

138) 유럽연합 중소기업과 공공분야의 클라우드 도입 확산을 위한 애플리케이션 유연성과 보안 강화

139) 클라우드 컴퓨팅과 포그 컴퓨팅이 혼합된 환경에서의 데이터 집중 애플리케이션

140) 지역적으로 분산된 페더레이티드 클라우드 환경에서의 데이터 집중 애플리케이션

141) 대용량 비디오 데이터 주석 시스템 (예: 자동차의 고급운전자보조시스템)

〈표 26〉 H2020 클라우드 컴퓨팅 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산(유로)	주관국가	주관기관 유형
	시작	종료	개월			
NECOS	'17.9	'21.2	42	1,494,906	스페인	대학
ATMOSPHERE	'17.11	'19.10	24	1,499,375	스페인	대학
UNICORN	'17.1	'19.12	36	3,415,487	독일	기업
ACTiCLOUD	'17.1	'19.12	36	4,733,533	그리스	연구소
COLA	'17.1	'19.6	30	2,975,081	영국	대학
DITAS	'17.1	'19.12	36	4,420,188	스페인	기업
MELODIC	'16.12	'19.11	36	4,890,224	노르웨이	대학
Cloud-LSVA)	'16.1	'18.12	36	4,604,431	스페인	연구소

⑤ OSS

H2020에서 진행하고 있는 과제들 중 OSS와 관련된 것으로 168개 과제가 확인되었다. 그 중에서 최근에 시작한 것들, 그리고 100만 달러 이상의 대형 연구를 중심으로 100개 과제를 검토하여 15개 예비 후보를 선발하였다. 검토에는 앞에서 밝힌 세 가지 기준을 적용하였다. 선발된 **과제**의 수행기간은 36개월과 48개월, 60개월로 다양하며, 연도별 예산은 모두 1,000만 달러가 넘는다.

- Open-source freeware for fiber optic communication and sensing simulations (SIMFREE)¹⁴²⁾
- ElasTest: an elastic platform for testing complex distributed large software systems (ELASTEST)¹⁴³⁾
- Developer-Centric Knowledge Mining from Large Open-Source Software Repositories (CROSSMINER)¹⁴⁴⁾
- Statistical multi-Omics UNDERstanding of Patient Samples (SOUND)¹⁴⁵⁾

142) 광 통신 및 센싱 시뮬레이션을 위한 오픈소스소프트웨어

143) 대형 소프트웨어 시스템을 위한 테스트 플랫폼

144) 소프트웨어 개발자의 오픈소스소프트웨어 컴포넌트 선택을 돕기 위해 대규모 오픈소스소프트웨어 리파지토리의 코드, 문서, 온라인 토론, 이슈 등을 데이터 마이닝

- Service Programing and Orchestration for Virtualized Software Networks (SONATA)¹⁴⁶⁾
- FREE and open source software tools for WATer resource management (FREEWAT)¹⁴⁷⁾
- Enabling Federated Cloud Networking (BEACON)¹⁴⁸⁾
- AppHub, the European Open Source Marketplace (AppHub)¹⁴⁹⁾

〈표 27〉 H2020 OSS 예비 후보의 수행 기간과 예산, 주관기관

과제	기간			예산(유로)	주관국가	주관기관 유형
	시작	종료	개월			
SIMFREE	'17.7	'19.6	24	134,462	폴란드	연구소
ELASTEST	'17.1	'19.12	36	4,270,187	스페인	대학
CROSSMINER	'17.1	'19.12	36	4,519,008	영국	기업
SOUND	'15.9	'18.8	36	2,953,063	독일	기업
SONATA	'15.7	'17.12	30	6,657,721	스페인	기업
FREEWAT	'15.4	'17.9	30	1,411,163	이탈리아	대학
BEACON	'15.2	'17.7	30	3,570,250	벨기에	연구소
AppHub	'15.1	'16.12	24	849,260	독일	연구소

145) 질병 이해와 새로운 처방 발견, 개인 맞춤 처방 실현을 위한 다수 유전자 정보의 통계 분석 툴

146) 네트워크 가상화 환경에 서비스 개발과 제공을 위한 툴

147) 물자원 관리를 위한 오픈소스소프트웨어

148) OpenNubula와 OpenStack 기반의 클라우드 페더레이션

149) 오픈소스소프트웨어 장터

제3절 후보 선정

1. 예비 후보 중 후보선정

후보 선정은 예비 후보들을 대상으로 각 기관별 및 분야별로 3~4명의 전문가들 의견을 종합해서 이루어졌다. 우선 DARPA와 IARPA는 다양한 영역의 기술이 포함되어 있어 최근 기술 기획 임무를 담당하고 있거나 담당할 경험이 있어 여러 기술 분야를 아우를 수 있는 전문가들로 구성했으며, 양 기관의 예비 후보를 함께 검토했다. NSF와 H2020은 5개 영역별로 과제 기획을 한 경험이 있거나 실제 해당 영역의 과제 책임자를 수행하고 있는 전문가로 검토위원을 선정했다. 이때 NSF와 H2020의 세부 영역이 같기 때문에 해당 영역의 검토위원이 양 기관의 예비 후보를 평가하도록 했다.

한편 전문가들은 아래와 같은 후보 선정 기준에 기초하여 예비 후보 중 적극추천과 배제희망 과제를 선정하도록 했으며, 두 가지에 해당하지 않는 과제는 중간으로 간주했다. 이때 후보 간 변별력 확보를 위해 적극 추천의 경우 +2점으로 하였고, 배제희망은 -1점, 중간은 0점을 주었다. +3점 이상 종합점수를 받은 과제를 후보로 선정하는 것을 원칙으로 했는데, 영역별 안배를 고려하여 +4점으로 한 경우도 있다.

〈표 28〉 후보선정 기준

- | |
|--|
| <ul style="list-style-type: none">• 소프트웨어 연구개발 과제로서 적절한가? (과제 내 소프트웨어의 비중 참고)• 기술 격차를 획기적 해소하는데 기여할 수 있는가?• 중장기 대형 과제로 추진할 필요가 있는가? (5년 이상 연 50억 원 이상 투자 필요) |
|--|

총 3,332개의 대상 과제 중에서 선발한 108개 예비 후보를 검토하였으며 그 중 21개를 후보로 선정했다. 예비 후보 대비 선정 비율은 19.44%로 대략 5개의 예비 후보 중 하나가 후보로 선정되었다. DARPA의 경우 가장 많은 4개의 후보가 선정됐으며, H2020의 빅데이터 과제는 모든 과제가 배제희망 의견이 나왔고 종합점수도 미달하여 하나도 선정되지 않았다. 선정된 21개의 후보 목록을 기관과 분야별로 구분하면 다음과 같다.

〈표 29〉 중장기 대형 R&D과제 발굴 과정을 통해 선정한 후보

후보 출처		후보
DARPA		<ul style="list-style-type: none"> • Active Interpretation of Disparate Alternatives (AIDA) • Data-Driven Discovery of Models (D3M) • Synergistic Discovery and Design (SD2) • Probabilistic Programming for Advancing Machine Learning (PPAML)
IARPA		<ul style="list-style-type: none"> • Active Interpretation of Disparate Alternatives (AIDA)
NSF	CPS	<ul style="list-style-type: none"> • Foundations of Resilient CybEr-Physical Systems (FORCES) • Correct-by-Design Control Software Synthesis for Highly Dynamic Systems
	빅데이터	<ul style="list-style-type: none"> • Big Data Modeling and Analysis with Depth and Scale • A Formal Foundation for Big Data Management
	HPC	<ul style="list-style-type: none"> • Compute on Data Path: Combating Data Movement in High Performance Computing • Hierarchical Tuning of Floating-Point Computations
	클라우드	<ul style="list-style-type: none"> • Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation • A Computing Cloud for Graphical Simulation
	OSS	<ul style="list-style-type: none"> • Chameleon: A Large-Scale, Reconfigurable Experimental Environment for Cloud Research • BONSAI: An Open Software Infrastructure for Parallel Autotuning of Computational Kernels
H2020	CPS	<ul style="list-style-type: none"> • SAFety and secURity by design for interconnected mixed-critical cyber-physical systems (SAFURE) • Integrated Modelling, Fault Management, Verification and Reliable Design Environment for Cyber-Physical Systems (IMMORTAL)
	빅데이터	해당사항 없음
	HPC	<ul style="list-style-type: none"> • European Exascale Processor Memory Node Design (ExaNoDe)
	클라우드	<ul style="list-style-type: none"> • Multi-cloud Execution-ware for Large-scale Optimized Data-Intensive Computing (MELODIC)
	OSS	<ul style="list-style-type: none"> • Enabling Federated Cloud Networking (BEACON) • Developer-Centric Knowledge Mining from Large Open-Source Software Repositories (CROSSMINER)

2. 절대 비교 대상 후보

한편 본 연구에서는 최종적으로 선정된 후보가 경쟁 후보들보다 상대적으로 우수한 것뿐만 아니라 절대적으로 우리나라에서 중장기 대형 과제로 추진할 가치가 있어야 한다. 그래서 절대 비교 대상으로 현재 국내에서 진행되고 있는 중장기 대형 과제 중 하나를 선정했다. 본 연구를 통해 선정된 최종 과제는 현재 진행되고 있는 중장기 대형 과제보다는 낫다는 평가를 받을 필요가 있는 것이다.

절대 비교 대상은 최종 선정을 위한 세 차례의 후보 검토 중 마지막 후보 검토에서만 사용되었다. 본 과제의 최종 후보 선정이 병렬로 진행되지 않고 누적적으로 이루어지기 때문에 1차와 2차의 검토 이후의 마지막 검토에서만 적용해도 최종 선정 후보와의 절대 비교가 가능하다.

2013년 이후 선정하여 추진하고 있는 소프트웨어 분야의 중장기 대형 과제는 총 6개이다. SW컴퓨팅산업원천기술개발사업 국가혁신형에 해당하는 엑소브레인(2013년)과 딥뷰(2014년), 빅데이터 처리용 고성능 컴퓨팅(2015년), 브레인 컴퓨터 인터페이스(2017년)가 있고, 인공지능 분야의 지능정보 플래그십 프로젝트(2016년)와 인공지능 국가전략 프로젝트(2017년)가 있다. 6개 과제 모두 연 기준 예산이 50억원이 넘고, 수행기관도 빅데이터 고성능 컴퓨팅(4년)을 제외하고 모두 5년 이상이다.

그런데 6개 중장기 대형 과제는 과제 단위에서 연구 목표와 내용을 제시하고 있는 경우도 있고 그렇지 않은 경우도 있다. 엑소브레인과 딥뷰, 빅데이터 고성능 컴퓨팅은 총괄과제가 별도로 있다. 브레인 컴퓨터 인터페이스와 지능정보 플래그십 프로젝트, 인공지능 국가전략 프로젝트는 총괄과제가 구분되어 있지만 1세부과제에 통합되어 있다. 6개 중장기 대형 과제들의 세부과제와 수행기관, 예산 등은 다음의 표와 같다.

〈표 30〉 현재 국내에서 수행중인 중장기 대형 소프트웨어 과제

과제	세부과제	수행기관	연간예산 (백만원)	비고
엑소브레인	총괄 세부과제 150)	ETRI	100	
	1세부과제 151)	ETRI	3,500	
	2세부과제 152)	솔트룩스	2,300	
	3세부과제 153)	KAIST	600	
	4세부과제 154)	ETRI	1,500	조기 종료

답류	총괄 세부과제 ¹⁵⁵⁾		ETRI	100	
	1세부과제 ¹⁵⁶⁾		ETRI	3,800	
	2세부과제 ¹⁵⁷⁾		광주과학기술원	2,000	
	3세부과제 ¹⁵⁸⁾		계명대학교	1,500	조기 종료
	4세부과제 ¹⁵⁹⁾		포항공대/ 서울대	600	
빅데이터 고성능 컴퓨팅	총괄 세부과제 ¹⁶⁰⁾		UNIST	3,000	
	1세부과제 ¹⁶¹⁾		알티베이스	2,400	
	2세부과제 ¹⁶²⁾		이지서티	1,600	
브레인 컴퓨터 인터페이스	총괄 및 1세부 과제 ¹⁶³⁾		KIST	3,000	
	2세부과제 ¹⁶⁴⁾		고려대	2,000	
지능정보 플러그십 프로젝트	총괄 및 1세부과제 ¹⁶⁵⁾		KETI	5,000	
	2세부과제 ¹⁶⁶⁾		KAIST	4,000	
	3세부과제 ¹⁶⁷⁾		경북대	3,000	
	4세부과제 ¹⁶⁸⁾		KAIST	3,000	
인공지능 국가전략 프로젝트	차세대 학습·추론	1세부과제 ¹⁶⁹⁾	UNIST	1,875	
		2세부과제 ¹⁷⁰⁾	KAIST	1,875	
	비디오 튜링 테스트	총괄 ¹⁷¹⁾ 및 1세부과제 ¹⁷²⁾	서울대	1,750	
		2세부과제 ¹⁷³⁾	KAIST	1,250	
		3세부과제 ¹⁷⁴⁾	코난테크놀로지	1,200	

주) 예산은 과제 선정 당시 금액

- 150) 사용자와 의사소통을 통한 지식공유 및 지능진화가 가능한 엑소브레인 SW 기술 개발
151) 휴먼 지식증강 서비스를 위한 지능진화형 WiseQA 플랫폼 기술 개발
152) WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발
153) 비구조적 기법 기반 인간모사형 자가학습 지능 원천기술 개발 / 컨텍스트 인지형 Deep-Symbolic 하이브리드 지능 원천 기술 개발 및 언어 지식 자원 구축
154) 자율지능형 지식/기기 협업 프레임워크 기술 개발
155) 대규모 실시간 영상 이해 기반의 시각지능 플랫폼 개발
156) 실시간 대규모 영상 데이터 이해·예측을 위한 고성능 비주얼 디스커버리 플랫폼 개발
157) 대규모 실시간 비디오 분석에 의한 전역적 다중 관심객체 추적 및 상황 예측 기술 개발
158) 환경 재난·재해의 조기 감지를 위한 다형 영상 및 복합 센서 데이터 기반의 분석·예측 기술 개발
159) 예지형 시각 지능 원천 기술 개발
160) 빅데이터 처리 고도화 핵심 기술개발 사업 총괄 및 고성능 컴퓨팅 기술을 활용한 성능 가속화 기술 개발
161) 분산환경 인메모리 기술 기반의 복합형 고속 스트림 빅데이터 처리 기술 개발
162) 빅데이터 환경에서 비식별화 기법을 이용한 개인정보보호 기술 개발
163) 생각만으로 실생활 기기 및 AR/VR 디바이스를 제어하는 비침습 BCI 통합 뇌인지컴퓨팅 SW 플랫폼 기술 개발
164) 딥러닝을 이용하여 사람의 의도를 인지하는 BCI 기반 뇌인지컴퓨팅 기술 개발
165) 자율지능 디지털 동반자 프레임워크 및 응용 연구

6개 중장기 대형 소프트웨어 과제들 중 비교 대상을 선정할 때 우선 최근에 선정된 과제를 선택할 필요가 있다. 시간에 따른 차이를 최소화하기 위함이다. 이 기준으로 보면 인공지능 국가전략 프로젝트, 브레인 컴퓨터 인터페이스, 지능정보 플래그십 프로젝트, 빅데이터 고성능 컴퓨팅, 딥뷰, 엑소브레인의 순서로 우선순위가 높다.

그런데 후보들과의 내용 비교를 위해 과제 전체의 내용을 파악할 수 있는 공개된 문서가 있어야 한다. 그래야 후보 검토에 필요한 객관적인 설명자료를 제공할 수 있다. 즉, 총괄 세부과제의 연구계획서나 제안요청서(RFP: Request for Proposal)를 확보할 수 있어야 한다. 인공지능 국가전략 프로젝트는 차세대 학습·추론과 비디오 튜링 테스트와 같이 서로 독립된 주제로 구성되어 있다. 실제 2개의 서브 사업단을 구성하였으며, 별도 총괄 세부과제가 없어서 과제의 전체의 목표와 내용을 설명하는 자료를 확보할 수 없었다. 브레인 컴퓨터 인터페이스의 경우에는 총괄세부과제 겸 1세부과제가 존재하였으나 관련 자료를 확보하지 못하였다. 한편 지능정보 플래그십 프로젝트의 총괄 및 1세부과제의 제안요청서를 확보할 수 있었다.

그래서 최종 후보 선정을 위한 기존 중장기 대형 절대 비교 후보로 지능정보 플래그십 프로젝트를 선정하였다. 지능정보 플래그십 프로젝트는 사용자를 이해하고 적절하게 도울 수 있는 적응형 학습 기계를 만드는 것을 최종 목표로 하고 있다. 175)

166) 자율지능 동반자를 위한 적응형 기계학습 기술 연구

167) 사용자의 의도와 맥락을 이해하는 지능형 인터랙션 기술 연구개발

168) 대화 상대의 감성 추론 및 판단이 가능한 감성 지능 기술 연구

169) 의사결정 이유를 설명할 수 있는 인간 수준의 학습·추론 프레임워크

170) 설명 가능한 인간 수준의 딥 기계학습 추론 프레임워크 개발

171) 인간 수준의 비디오 이해 지능 및 검증 기술 개발

172) 비디오 튜링 테스트를 통과할 수준의 비디오 스토리 이해 기반의 질의응답 기술 개발

173) 비디오 이해를 위한 이벤트-상황 지식체계 학습 및 이벤트 인식/관계 추론 기술 개발

174) 비디오 이해를 위한 데이터 수집 및 보정 자동화 시스템 개발

175) 자세한 내용은 <부록 23> 참조

제4절 최종 후보 선정

최종 후보는 10명의 후보 검토 위원이 22개 후보를 검토한 결과를 종합하여 이루어졌다. 후보 검토 위원 각자 평가항목의 중요도를 설정한 후 복수의 후보 쌍을 대상으로 상대비교를 하면 개별 검토 위원의 각 후보별 점수가 계산된다. 후보별 평가 점수는 후보 검토 위원의 평가값을 평균하여 산출하였다. 후보별 평가 점수의 계산식은 아래와 같다.

$$\text{후보별 평가 점수} = \text{AVERAGE} \left(\sum \text{평가항목 } i \text{의 중요도} \times \text{평가항목 } i \text{의 평가값} \right)$$

전문가들은 후보 검토에 앞서 중장기 대형 과제 발굴에 필요한 아래와 같은 2가지 평가 항목의 중요도에 대한 가중치를 설정했다. 두 가지 평가 항목에 ~~서~~ 대한 중요도를 총점을 100으로 하여 검토 위원 각자 설정하였다. 즉, 검토 위원 각자 기술 격차 해소 가능성과 우리나라의 연구진의 기술개발 성공 가능성의 합이 100이 되도록 두 평가항목에 점수를 배분하였다. 이렇게 설정한 중요도는 세 차례에 걸친 후보 검토의 점수 산정에 모두 적용됐다.

<표 31> 최종 후보 선정 기준

- 기술 선도국과의 기술 격차를 획기적으로 해소할 수 있다 (기술 격차 해소 가능성)
- 우리나라 연구진이 개발에 성공할 것이다 (기술개발 성공 가능성)

후보 검토는 앞서 밝힌 바와 같이 세 차례로 분할되어 진행되었는데, 병렬적으로 진행되지 않았다. 제1차 후보 검토는 DARPA와 IARPA 기준 후보들만을 대상으로 하였다. 제2차 후보 검토는 제1차 후보 검토로부터 생존한 후보와 새로운 후보를 추가하여 이루어졌는데, 새롭게 추가한 후보로는 NSF의 CPS와 빅데이터, HPC, 클라우드 컴퓨팅 과제 기준의 후보들이었다. 제3차 후보 검토 역시 제2차 후보 검토로부터 생존한 후보와 새로운 후보를 추가하여 제2차 후보 검토가 이루어졌다. 새롭게 추가된 후보는 NSF의 오픈소스소프트웨어(OSS) 과제와 EU H2020의 과제 기준의 후보들이었다. 그리고 제3차 후보 검토로부터 생존한 후보를 최종 후보로 결정하였다.

1. 1차 후보 검토

1차 후보 검토는 검토 위원들이 새로운 검토 방식에 익숙해질 수 있도록 DARPA와 IARPA 과제 기준 5개 후보만으로 제한하였다. 검토 대상인 5개 후보는 다음과 같다.

- Active Interpretation of Disparate Alternatives (AIDA)
- Synergistic Discovery and Design (SD2)
- Data-Driven Discovery of Models (D3M)
- Probabilistic Programming for Advancing Machine Learning (PPAML)
- Deep Intermodal Video Analytics (DIVA)

10명 검토 위원들 중 제출 마감 시간에 맞추어 검토 의견을 제출한 7명의 위원의 검토 결과는 다음과 같다. 5개의 후보 중 DIVA를 1위로 선정하였다. 그 다음으로는 AIDA와 D3M, SD2, PPAML의 순서로 높은 점수를 부여하였다. DIVA는 4명의 검토 위원이 최고로 평가했지만 2명은 최저점을 부여했다. AIDA와 D3M에 대해 최저점을 주지 않고 최고 높은 점수만 부여한 위원이 1명씩 있었다. SD2는 최고점을 부여한 위원이 있었지만 그 보다 많은 수가 최저점을 부여하였다. PPAML은 최저점을 부여한 위원만 있었다. 종합 순위를 고려하여 1위부터 4위까지인 DIVA와 AIDA, D3M, SD2를 2차 후보 검토의 후보로 선정하였다.

〈표 32〉 1차 후보 검토 결과

구분	사업명	최고값 부여(명)	최저값 부여(명)	점수	순위
DARPA	AIDA	1	0	1.151	2
	SD2	1	3	0.959	4
	D3M	1	0	1.098	3
	PPAML	0	2	0.957	5
IAPRA	DIVA	4	2	1.200	1

2. 2차 후보 검토

1차 후보 검토에서 생존한 4개 후보에 NSF의 CPS와 빅데이터, HPC, 클라우드에서 선정한 8개의 후보를 함께 검토하였다. 10명의 검토 위원에게 의뢰하였는데, 마감 시간에 맞추어 응답한 6명의 의견을 종합하였다. 검토결과 최고점은 1차와 동일하게 IARPA의 DIVA과제가 받았다. 점수 순위 기준으로 4위까지 선정하였는데, DARPA 1개, IARPA 1개, NSF 2개 후보를 3차 검토 후보로 선정했다.

- [후보01] Foundations of Resilient CybEr-Physical Systems (FORCES)
- [후보02] Correct-by-Design Control Software Synthesis for Highly Dynamic Systems
- [후보03] Active Interpretation of Disparate Alternatives (AIDA)
- [후보04] Big Data Modeling and Analysis with Depth and Scale
- [후보05] A Formal Foundation for Big Data Management
- [후보06] Synergistic Discovery and Design (SD2)
- [후보07] Compute on Data Path: Combating Data Movement in High Performance Computing
- [후보08] Hierarchical Tuning of Floating-Point Computations
- [후보09] Data-Driven Discovery of Models (D3M)
- [후보10] Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation
- [후보11] A Computing Cloud for Graphical Simulation
- [후보12] Deep Intermodal Video Analytics (DIVA)

〈표 33〉 2차 후보 검토에서의 전문가 의견 종합

구분	사업/과제명	최고값 부여(명)	최고값 부여(명)	점수	순위
DARPA	[후보03] AIDA	1.5 ^주	0	1.268	3
	[후보06] SD2	0	1	1.070	10
	[후보09] D3M	0	0	1.102	8
IAPRA	[후보12] DIVA	1	0	1.336	1
NSF	[후보01] FORCES	0	1	0.955	11
	[후보02]	1	0	1.283	2
	[후보04]	1	1	1.202	5
	[후보05]	0.5 ^주	0	1.226	4
	[후보07]	0	0	1.119	7
	[후보08]	0	3	0.840	12
	[후보10]	1	0	1.079	9
	[후보11]	0	0	1.178	6

주) 1명의 전문가가 2개의 후보에 최고점을 부여하여 각 후보에 0.5명씩 부여

3. 3차 후보 검토

2차 후보 검토에서 생존한 4개에 NSF의 Open Source SW와 EU의 과제에서 각각 선택한 2개와 6개의 후보를 함께 검토하였다. 한편 앞 절에서 설명했듯이 기존 국내에서 수행되고 있는 과제와의 절대 비교를 위해 지능정보 플래그십 프로젝트에서 수행중인 1개 과제, [후보1]을 추가해서 검토했다.

- [후보01] 자율지능 디지털 동반자 프레임워크 및 응용 연구개발
- [후보02] Correct-by-Design Control Software Synthesis for Highly Dynamic Systems
- [후보03] SAFety and secURity by design for interconnected mixed-critical cyber-physical systems (SAFURE)
- [후보04] Integrated Modelling, Fault Management, Verification and Reliable Design Environment for Cyber-Physical Systems (IMMORTAL)
- [후보05] European Exascale Processor Memory Node Design (ExaNoDe)
- [후보06] Multi-cloud Execution-ware for Large-scale Optimized Data-Intensive Computing (MELODIC)

- [후보07] Chameleon: A Large-Scale, Reconfigurable Experimental Environment for Cloud Research
- [후보08] Enabling Federated Cloud Networking (BEACON)
- [후보09] BONSAI: An Open Software Infrastructure for Parallel Autotuning of Computational Kernels
- [후보10] Developer-Centric Knowledge Mining from Large Open-Source Software Repositories (CROSSMINER)
- [후보11] Active Interpretation of Disparate Alternatives (AIDA)
- [후보12] A Formal Foundation for Big Data Management
- [후보13] Deep Intermodal Video Analytics (DIVA)

1차와 2차 후보 검토와 마찬가지로 10명의 전문가에게 의뢰하였다. 마감 시간에 맞추어 10명 모두 검토 의견을 제시하였다. 절대 비교 기준(후보1)보다 4개 과제가 높은 점수를 받아 이들 과제를 최종 후보로 선정했다.

〈표 34〉 3차 후보 검토에서의 전문가 의견 종합

구분	사업/과제명	최고값 부여(명)	최저값 부여(명)	점수	순위
DARPA	[후보11] AIDA	0	0	1.067	1
IAPRA	[후보13] DIVA	2	0	1.040	3
NSF	[후보2] Correct-by-Design Control SW Synthesis	1	1	0.934	7
	[후보7] Chameleon	0	3	0.886	11
	[후보9] BONSAI	0	0	1.029	4
	[후보12] A Formal Foundation of BD Management	0	1	0.898	10
EU	[후보3] SAFURE	1	1	0.980	5
	[후보4] IMMORTAL	0	2	0.874	12
	[후보5] ExaNoDe	2	2	0.952	6
	[후보6] MELODIC	0	0	0.934	7
	[후보8] BEACON	0	0	0.930	9
	[후보10] CROSSMINER	2	0	1.044	2
기존 과제	[후보1] 자율지능 디지털 동반자	1	0	1.010	열외

제6장 최종 후보별 기획방향

본 연구에서는 기술 선도국에서 현재 진행하고 있는 연구개발과제들 중 우리나라의 기술 격차 해소 내지 역전에 의미 있는 것들을 예비 후보와 후보, 최종 후보의 3단계 선발 및 선정 과정을 통해 도출하였다. DARPA의 AIDA와 IARPA의 DIVA, 미국 NSF의 BONSAI, EU H2020의 CROSSMINER의 4가지 과제가 그것들이다. 물론 이들 4개 최종 선정 후보들과 동일한 연구개발과제를 그대로 기획할 수는 없다. 우리나라의 현실을 반영하여 해당 후보들을 수정해야 한다.

한편, 최종 후보들이 실제 연구개발과제로 기획되기 위해서는 수요조사에서 제기된 과제들과의 비교가 가능해야 한다. 그래야 수요조사의 제안 과제들과 비교해서 우수한 것을 선정할 수 있기 때문이다. 따라서 최종 선정된 각 후보들에 대해 수요조사 제안 과제들과 동등한 수준의 내용이 기획되어야 한다. 기술수요조사에서는 제안자가 연구 과제의 개발 목표, 개발 내용, 국내외 기술개발 동향, 지원 필요성 및 기대효과 등을 기술하도록 되어 있다.¹⁷⁶⁾ 따라서 본 연구에서는 선정된 각 후보의 연구개발 목표 및 내용, 국내외 동향, 정부 지원 필요성 및 기대효과를 작성하는데 필요한 일차적인 조사 내용을 정리하고 향후 과제를 기획할 때 고려해야 하는 사항을 제시하고자 한다.

제1절 멀티소스멀티모달 기계학습 및 다중 가설 자동 생성

1. DARPA AIDA 과제

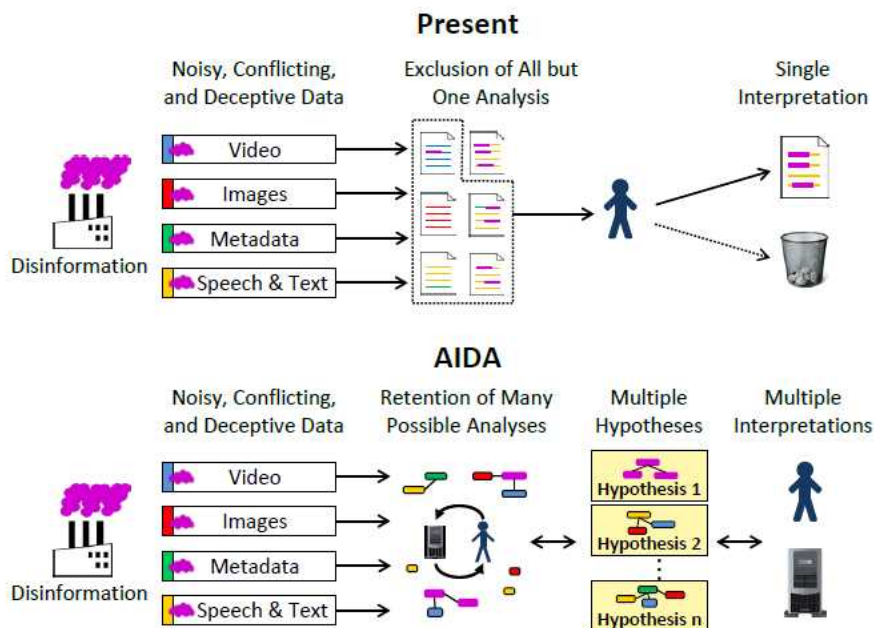
DARPA Active Interpretation of Disparate Alternatives (AIDA)¹⁷⁷⁾는 정보 분석가의 분석 역량을 강화시키기 위한 사업이다. 일반적으로 정보 분석가는 시간적 제약과 인지적 한계로 인해 제한된 종류와 수의 데이터를 분석하여 한 가지 가설에 대해 검증한다. 그리고 이와 같은 제약은 정보 분석가가 고정 관념을 벗어나기 어렵게 만든다. 특히, 지금까지 유지된 추세가 전환되는 시점에 그 전조를 놓치는 실수를 범할 수 있다. 예를 들면, 2014년 초 미국 정부의 담당자들은 러시아의 크림반도 진군 가능성을 확인할 수 있는 여러 정보를 입수하였음에도 불구하고 러시아가 군사적 행동을 취할 움직임은 없다고 판단하였다.¹⁷⁸⁾

176) 수요조사 제안 중 비교 평가를 통해 선정되어 과제뱅크에 등록될 때 연구개발 목표 및 내용, 정부 지원 필요성, 활용 분야(활용 제품 및 시장) 등을 기재함

177) <https://www.darpa.mil/program/active-interpretation-of-disparate-alternatives>

AIDA는 위에서 제기한 근본적 한계의 해소를 목적으로 하고 있다. 이를 위해 먼저 특정 사건(event)이나 상황(situation), 추세(trend)에 대한 복수 가설을 자동으로 생성하여 정보 분석가에게 제공하고자 한다. 자동 생성된 복수의 가설을 전달 받은 정보 분석가는 마찬가지로 자동으로 평가된 각 가설의 확신도(confidence)와 자신의 전문가적 견해를 종합하여 최종 판단을 내리게 된다. 이때 가설을 자동으로 생성하기 위해 기계 학습을 사용한다. 학습에 이용할 데이터는 텍스트, 이미지, 영상 등 제한을 두지 않고 있기 때문에 일명 멀티모달(multi-modal) 학습을 전제로 한다. 또, 지리적 위치, 시간 등과 같은 메타 데이터까지 결합할 것이다. AIDA 과제의 혁신성은 아래의 그림을 통해 확인할 수 있다.

<그림 7> DARPA AIDA사업 전과 후의 정보분석의 차이



출처: <https://www.darpa.mil/attachments/AIDA%20Proposers%20Day%20v7.pdf>

AIDA는 빅데이터 분석에 해당되면서 동시에 인공지능에도 해당된다.¹⁷⁹⁾ DARPA는 이 사업의 세부 과제(TA: Technical Area)를 다음과 같이 5가지로 구성했다. 첫째, 각 데이터 출처로부터 지식 요소(knowledge element)¹⁸⁰⁾를 추출하고 동 사업에서 고안한 공통 의미 표현(common semantic representation) 모형의 적용 방법을 개발한다. 이때,

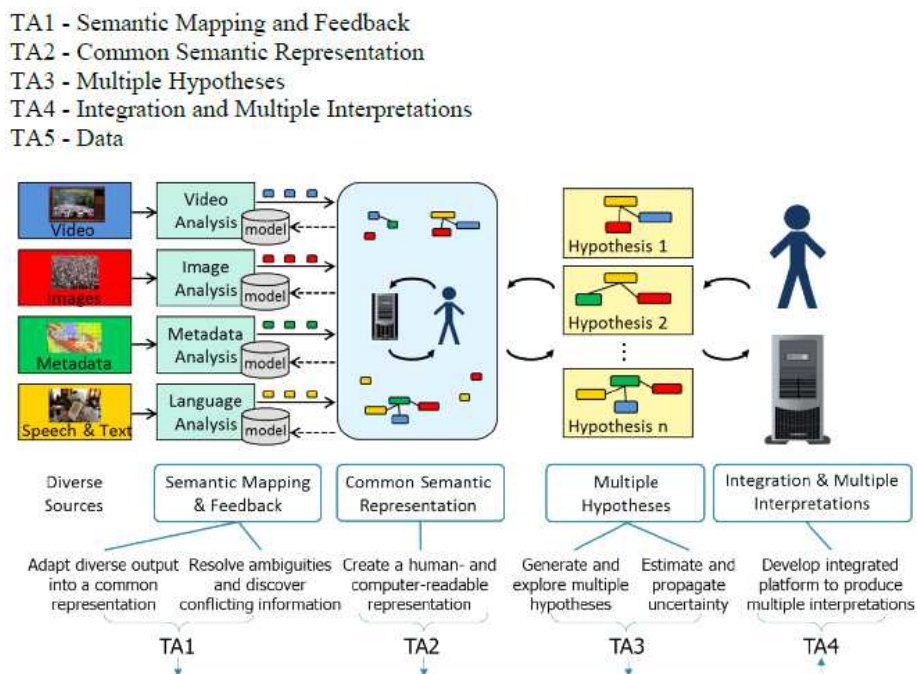
178) AIDA의 러시아 크림반도 침공 분석 시나리오는 <부록 25> 참조

179) DARPA는 AIDA의 색인으로 Data와 Analytics, Language, Imagery, Autonomy, AI를 설정

180) 사건 또는 하위 사건, 행동, 엔터티(entity), 관계, 장소, 시간, 긍정 또는 부정 감정 등에 관한 구조화된 표현(structured representation)

러시아어 트위터처럼 외국어를 번역하여 지식 요소를 추출할 수도 있어야 한다. 둘째, 공통 의미 표현 모형을 개발하고 이 공통 모형을 통해 표현된 지식 요소를 지식 기반(knowledge base)¹⁸¹⁾으로 통합한다. 이때, 엔터티(entity)의 의미적 모호성까지 제거할 수 있어야 한다. 셋째, 공통 의미 표현 모형에 기초하여 내적 일관성과 외적 판별성을 갖는 복수 가설을 추출하고, 각 가설의 확신도를 수치화하여 제시한다. 넷째, 위의 세 가지 기술 분야에서 개발된 것들을 바탕으로 프로토타입(prototype)을 개발한다. 사용자가 지식 기반 탐색, 가설 표시, 쿼리 형태의 검색, 지식 요소 수정 등과 같은 시스템과의 상호작용이 가능해야 한다. 마지막으로 AIDA 개발에 사용할 데이터를 마련하는데 필요한 기술을 개발하는 것이다. 데이터는 4개의 시나리오를 위한 것인데, 시나리오로는 국제 분쟁과 자연 재해, 국제 행사에서의 폭력 또는 테러, 시위를 설정하고 있다. 특히, 4개의 시나리오 중 하나는 학습용 시나리오로서 동 사업의 킥오프 미팅(kick-off meeting) 때 주석(annotation)이 되어 있는 기본 데이터가 제공될 것이다. AIDA의 5가지 세부 과제의 내용과 관계는 아래의 그림과 표에서 보는 바와 같다.

〈그림 8〉 DARPA AIDA사업의 구성



출처: <https://www.darpa.mil/attachments/AIDA%20Proposers%20Day%20v7.pdf>

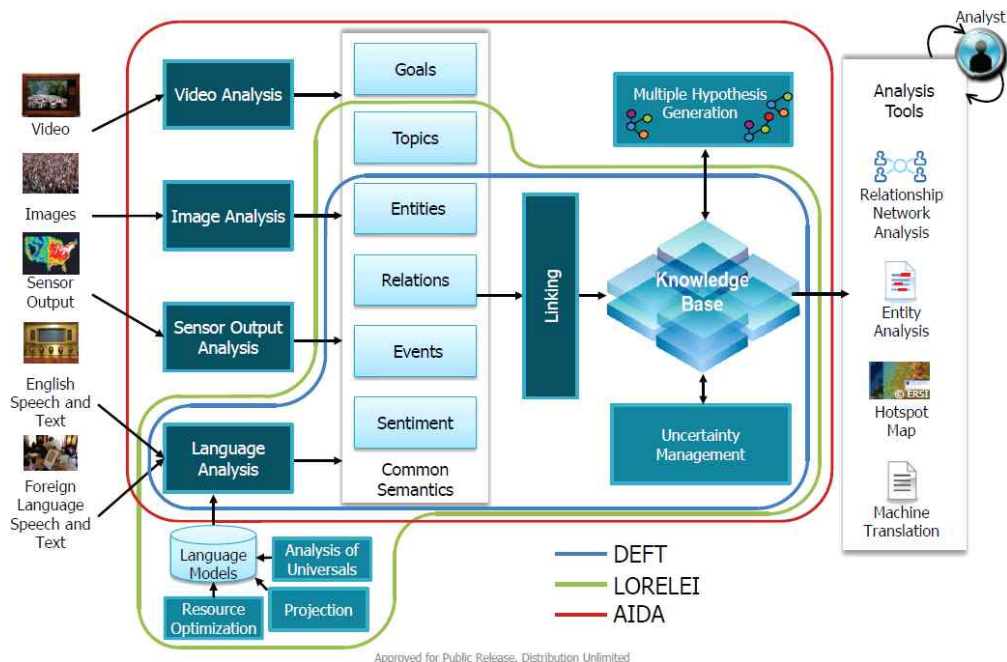
181) 여러 지식 요소가 저장되어 있는 논리적 또는 물리적 공간

〈표 35〉 데이터 세부과제를 제외한 4개 세부 과제의 입출력 관계

From	입력	TA	출력	To
미디어	가공전 데이터	TA1	지식 요소	TA2
TA3 TA4	가설			
TA1	지식 요소	TA2	지식요소	TA3 TA4
TA3	가설			
TA4	지식 요소 및 가설			
TA2	지식 요소	TA3	가설	TA1 TA2 TA4
TA4	지식 요소 및 가설			
TA2 TA3 사용자	지식 요소	TA4	지식 요소 및 가설	사용자 TA1 TA2 TA3
	가설			
	지식 요소 및 가설			

한편, AIDA는 DARPA의 여러 과제들과 연관이 깊다. 그 중에서 LORELEI¹⁸²⁾과 DEF T¹⁸³⁾의 관계는 아래 그림과 같다. 그림에서 보는 바와 같이 AIDA는 기존의 두 과제에 비해 멀티모달 학습과 복수 가설 생성이라는 측면에서 차이를 보인다.

〈그림 9〉 AIDA와 LORELEI, DEFT의 관계



출처 <https://www.nap.edu/read/24900/chapter/6>

182) Low Resource Languages for Emergent Incidents (2015년~)

183) Deep Exploration and Filtering of Text (2012-2017년)

AIDA는 총 54개월(4년6개월) 동안 진행할 예정이며, 2017년과 2018년의 예산으로 각각 550만 달러와 1,730만 달러가 책정되어 있다. 이를 위해 2017년 6월6일까지 제안서 접수가 이루어졌다. 지금까지 어떤 기관의 어떤 연구자가 각 세부과제를 담당하고 있는지 구체적으로 밝혀진 바와 매우 적지만, 지금까지 확인된 것을 요약하면 아래와 같다.

남가주대학(USC) 정보과학원(Information Sciences Institute) 중심의 대학 컨소시엄이 AIDA의 첫 번째 세부과제(TA1)를 수주하였다. 지원 기간은 54개월이고 지원 금액은 약 1,500만 달러이다. 그런데 1,500만 달러 중 233만 달러는 컬럼비아대학팀에 할당되었다.¹⁸⁴⁾ 이 팀에는 Kathleen McKeown(H-Index: 26)이 참여하고 있는데, 웹상의 대용량 데이터 추상 요약(abstract summarization)에 관한 여러 논문을 발표하였다. 2016년 IJCAI(International Joint Conference on Artificial Intelligence)에서 “Extractive and Abstractive Event Summarization over Streaming Web Text”를 발표하였다.¹⁸⁵⁾ 한편, Rensselaer Polytechnic Institut의 Heng Ji는 DEFT와 LORELEI에 이어 AIDA에도 참여하고 있다.¹⁸⁶⁾

2. 국내 관련 연구과제

인공지능의 분야별 연구가 활발하게 진행되고 있다. 언어지능의 엑소브레인 과제, 시각지능의 딥뷰 과제가 대표적이다. 엑소브레인은 “인간의 지적 노동을 보조할 수 있는 언어처리 분야의 AI 기술개발”이란 과제 개요에서 보는 바와 같이 언어지능 분야의 연구과제이다. 특히, “전문직 종사자와 지식 소통이 가능”하기 위해 ① 자연어 분석 기술 ② 지식 학습 및 축적 기술 ③ 자연어 질의응답 기술”로 세부과제가 구성되어 있다.¹⁸⁷⁾ 딥뷰는 “대규모 이미지/동영상을 실시간으로 수집하여 도심공간의 다차원/시계열 변화를 이해하고, 환경 재난·재해·위험 예측을 실시간으로 수행하는 시각 지능 플랫폼 기술 개발”이다.¹⁸⁸⁾ 그러나 이들 과제들은 모두 멀티모달 학습을 전제로 하고 있지 않다. 즉, 각각 언어지능과 시각지능 분야의 핵심 도전과제들을 해결하는 것을 목표로 하고 있다.

184) http://engineering.columbia.edu/web/newsletter/fall_2017/spot_news_roundup

185) <https://www.ijcai.org/Proceedings/16/Papers/575.pdf>

186) <http://nlp.cs.rpi.edu/hengji/grants.html>

187) <http://exobrain.kr/business/index.do>

188) NTIS에서 검색한 딥뷰 총괄과제 요약서 참고

한편, 2016년 말부터 멀티모달 학습과 관련된 중대형 과제들이 시작되었다. 예를 들면, 2016년 말에 시작한 지능정보 플래그십 프로젝트의 총괄 목표는 “사용자를 이해하고 적절한 도움을 주는 동반자 관점에서, 다양한 입력을 수집하여 학습하는 적응형 기계학습(Adaptive Machine Learning) 기반 자율지능 디지털 동반자 기술 개발”이다.¹⁸⁹⁾ 이때, “다양한 입력”에는 정형 데이터뿐만 아니라 비정형 데이터까지 모두 포함되고, 비정형 데이터에는 음성과 텍스트뿐만 아니라 제스처 등의 시각 정보까지 포함된다. 또, 2017년부터 시작한 인공지능 국가전략 프로젝트는 2개의 과제로 구성되어 있는데, 그 중 하나인 ‘비디오 튜링 테스트(Video Turing Test): 비디오 이해 지능 기술 개발 및 비디오 이해도에 대한 튜링 테스트 체계 개발’에도 멀티모달 학습이 포함되어 있다.¹⁹⁰⁾ 비디오의 이야기 전개를 이해하기 위해 비디오에 포함되어 있는 영상과 소리, 텍스트를 동시에 학습한다. 그러나 위의 두 중대형 과제 모두 단일 데이터 출처를 이용하고 있다. 지능정보 플래그십 프로젝트의 자율 동반자 과제는 사용자 데이터만을 대상으로 하고, 인공지능 국가전략 프로젝트의 비디오 튜링 테스트도 비디오 데이터만을 대상으로 한다. 그에 반해 AIDA는 다양한 출처의 다양한 형식의 데이터를 통합하여 학습하는 것을 전제로 한다. 그리고 두 프로젝트 모두 자동 가설 생성을 연구내용으로 삼고 있지 않았다.

한편 국내에서도 이중 빅데이터 통합 분석에 관한 연구과제들이 대학을 중심으로 진행되고 있다. 모두 차세대정보컴퓨팅기술개발사업 과제인데, 과제 규모가 몇 천 만원 수준으로 소규모이고, 개인 연구이자 기초연구이다. 각 연구과제의 제목과 책임자, 기간, 금액은 아래와 같다.

- 이중 빅데이터 통합 분석을 위한 메타러닝 모델 및 응용 연구 (동국대 김준태, ' 17.11.1.~' 20.12.31., 4,000만원, 차세대정보컴퓨팅기술개발사업)
- 이중 빅데이터 통합 분석 메타러닝 소프트웨어 개발 ((주)비아이매트릭스¹⁹¹⁾ 황선민, ' 17.11.1.~' 20.12.31., 2,625만원, 차세대정보컴퓨팅기술개발사업)
- 딥러닝 기법을 이용한 이중 빅데이터 통합 처리 모델 개발 (중앙대 임창원, ' 17.11.1.~' 20.12.31., 3,500만원, 차세대정보컴퓨팅기술개발사업)

189) 지능정보 플래그십 프로젝트 총괄과제 과제제안요청서(RFP) 참고

190) 다른 하나는 ‘의사결정 이유가 설명 가능한 휴먼 레벨 학습·추론 프레임워크’임

191) 비즈니스 인텔리전스(BI: Business Intelligence) 전문기업임 (<http://www.bimatrix.co.kr>)

참고로 ‘멀티모달’ 과 ‘딥러닝’ 을 모두 키워드로 갖는 국내 논문은 매우 적다. DBpia 검색에서는 총 6건이 검색되었는데, 음악 장르 분류, 제스처 인식, 동영상 분류 등 단일 출처의 멀티미디어 데이터를 학습 대상으로 하는 것들이다. AIDA에서와 같이 텍스트 데이터 등까지 학습에 포함하는 것이나 가설을 자동으로 설정하는 것은 내용으로 하고 있지 않다. 대표적인 논문의 제목과 저자, 발표처, 발표연도는 아래와 같다.

- 멀티모달 딥러닝을 이용한 음악 장르 분류 (김선훈, 김대식, 서봉원, 한국HCI학회 학술대회, 2016)
- 멀티모달 데이터를 활용한 딥러닝 기반 핸드 제스처 인식 알고리즘 (최재성, 이형민, 이상윤, 대한전자공학회 학술대회, 2017)
- 텐서플로우로 구현한 효과적인 구조의 멀티모달 딥러닝 동영상 분류 (한국정보과학회 학술대회, 2016)

3. 향후 기획 방향

멀티모달 학습과 자동 가설 생성을 기준으로 보면 DARPA의 AIDA와 우리나라의 관련 연구과제와는 차이가 있다. 우리나라의 멀티모달 학습에 관한 연구과제들은 2016년 이후부터 시작되었으며 대부분 멀티미디어 데이터의 학습에 치중해 있고, 자동 가설 생성을 목표로 한 연구과제는 없다. 이 차이를 그대로 두면 몇 년 후 우리나라와 기술 선도국과의 기술 격차는 더 크게 벌어질 것으로 예상된다. 따라서 지금 이 시점에 멀티모달 학습과 자동 가설 생성을 연구 내용으로 하는 연구과제의 기획을 적극적으로 검토해야 한다.

만약 우리나라에서 멀티모달 학습 및 자동 가설 생성 연구과제를 기획한다면 빅데이터와 인공지능 분야를 모두 포괄하는 차세대 연구과제가 될 것인데, 다음의 같은 점을 고려하여 추진할 필요가 있다. 첫째, 언어 선택 또는 범위의 문제이다. AIDA는 영어된 데이터뿐만 사건 일어난 현지의 언어로 된 텍스트의 번역까지 포함하고 있다. 우리나라의 입장에서는 한글 데이터만 대상으로 할 것인가와 한글 외 영어 등 외국어 데이터를 포함할 것인가 중 선택해야 한다. 그리고 언어 선택은 시나리오와 연결된다. 한글만을 대상으로 한다면 아무래도 국내와 관련된 시나리오를 선택해야 한다. 예를 들면, 국내에서 발생하는 지진 등의 자연재해나 사건사고, 국내에서 개최되는 행사를 대

상으로 할 수 있다. 반면 영어 등의 외국어까지 확대하게 되면 AIDA와 마찬가지로 국제 분쟁이나 북한 미사일 발사 가능성과 같은 시나리오를 검증할 수도 있을 것이다.

둘째, 멀티모달 학습 및 자동 가설 생성 연구과제가 성공하기 위해서는 무엇보다 국내 연구자들 사이의 교류와 협력이 필수적이다. 현재 개발이 이루어지고 있는 기술 수준을 뛰어 넘는 연구 목표를 설정하고 있기 때문에 우리나라의 언어지능과 시각지능 각각의 기술발전뿐만 아니라 멀티미디어 데이터 학습 분야의 기술발전까지 모두 종합해야 한다. 언어지능과 시각지능의 중장기 대형과제인 엑소브레인과 딥뷰, 멀티모달 학습을 표방하고 있는 지능정보 플래그십 프로젝트의 자율 동반자 과제와 인공지능 국가전략 프로젝트의 비디오 튜링 테스트, 차세대정보컴퓨팅기술개발사업의 이중 빅데이터 통합 분석 과제들에 참여하는 국내 연구자들이 새로운 과제에 직접 참여하는 것도 한 방법이다.

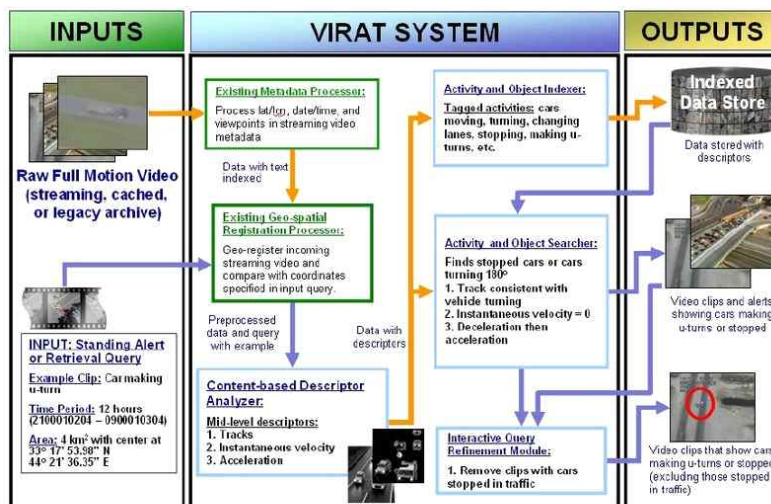
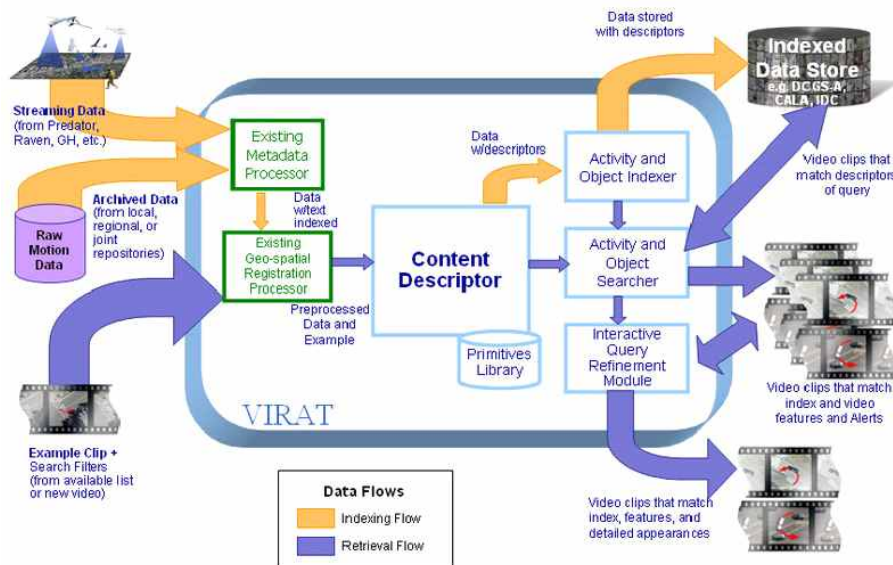
셋째, 해외 전문가들과의 협업이 중요하다. 만약 언어의 범위를 영어 등 외국어로 확대한다면 그 중요성을 더 커질 것이다. AIDA 사업에 참여하고 있는 미국 대학의 연구진들과 협력 방안을 적극적으로 모색해야 한다. 협력 대상의 예로는 남가주 대학이나 컬럼비아대학의 연구팀 또는 Rensselaer Polytechnic Institut의 Heng Ji 등이 될 수 있다. 연구 협력 방안에는 석박사 과정 학생들의 해외 파견 등도 고려해 볼 수 있다.

제2절 다중카메라 환경에서의 실시간 행위 인식 및 탐지

1. IARPA의 DIVA와 DARPA의 VIRAT

본 연구에서는 IARPA의 DIVA를 대상으로 후보 선정이 이루어졌다. 그러나 DIVA와 유사한 연구개발과제는 미국 연방정부의 여러 기관에서 기획되어 진행되어 왔다. 2009년부터 시작한 DARPA의 VIRAT¹⁹²⁾도 그 중 하나이다.¹⁹³⁾

<그림 10> DARPA VIRAT과제의 개념과 운영사례



출처: <https://en.wikipedia.org/wiki/VIRAT>

192) Video and Image Retrieval and Analysis Tool

193) <https://www.nap.edu/read/24900/chapter/4>

VIRAT은 특정 행동의 탐지를 목표로 하고 있다. 예를 들면, 화면에 등장하는 인물이 여러 행동을 하는 중에 땅을 파거나 물건을 집으면 기계가 인지하는 것이다. 개인의 행동뿐만 아니라 한 사람이 다른 사람 따라가기 등과 같은 사람과 사람 사이의 상호작용과 차량 탑승이나 문 열기 등과 같은 사람과 사물 사이의 상호작용에도 관심을 두고 있다. 한편 이 과제는 이미지와 영상 데이터의 인덱싱 방법과 검색 툴의 개발 뿐만 아니라 비디오 데이터셋 구축도 포함하고 있다. 비디오 데이터셋이 공개되었고, 2012년에 버전2.0¹⁹⁴⁾으로 업데이트되었다.

한편 IARPA는 Deep Intermodal Video Analytics (DIVA)사업¹⁹⁵⁾을 기획했다. 과제의 목표는 CCTV 등과 같은 복수의 스트리밍 카메라로부터 입수하는 영상 정보에서 객체와 행위를 자동으로 인식하고 탐지하는 것이다. 이때 하나 또는 여러 카메라에서 촬영한 동일 객체나 동일 행동을 하나로 인식하는 것도 포함한다. 이 과제가 성공하게 되면 정부 시설 대상의 테러 행위나 공공장소에서의 계획된 테러를 실시간으로 분석하여 보안 책임자에게 통보할 수 있을 것으로 기대된다. 또, 테러 이후에 대용량 비디오의 분석을 지금보다 더 빠르게 수행할 수 있을 것이다.

카메라는 기본적으로 지상에 설치된 것을 대상으로 하는데, 인도나 도로의 고정식 카메라, 실내나 건물 외곽에 설치된 CCTV나 웹캠이 여기에 해당된다. 그런데 이 과제에서 대상으로 삼고 있는 카메라의 범위는 휴대용 카메라나 휴대폰 카메라, 웨어러블 카메라로 확장할 것이다. 단, 3D가 아닌 2D 이미지 촬영 기능을 보유한 것으로 한정한다. 드론이나 인공위성 촬영 영상은 이 과제의 분석 대상이 아니다. 또 건물 주변과 같이 제한된 지역에 설치된 복수 카메라를 대상으로 할 뿐 넓은 지역을 대상으로 하지 않는다. 즉, 카메라의 화면이 서로 겹쳐서 상대적 위치를 확인할 수 있어야 한다. 음향은 분석 대상이 아니며, 하드웨어나 센서 개발도 연구 내용에 포함되지 않는다.

DIVA는 크게 행위 탐지와 대상 탐지로 구성되는데, 행위 탐지가 핵심 기능이다. 사람이나 사물 등의 대상 탐지는 행위 탐지를 보조하는 수단이다. DIVA에서 관심을 두고 있는 행위의 예로는 다음과 같다. 돌 등을 던지는 사람, 백팩(backpack) 등을 버리는 사람, 차량을 타거나 내리는 사람, 차량에서 물건을 내리는 사람, 무기를 휴대한 사람들, 물건을 주고받는 사람, 여러 명이 회합하거나 대화를 나누는 상황, 시위대, 휴대용품의 사용 등이 대표적이다.

194) <http://www.viratdata.org/>

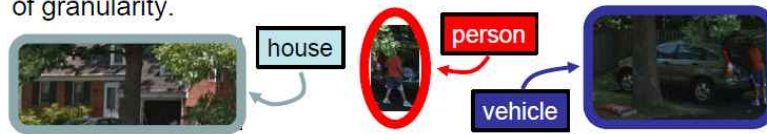
195) <https://www.iarpa.gov/index.php/research-programs/diva>

<그림 11> IARPA DIVA과제의 연구 목표

- Two main goals are evaluated:
 - (a) **Activity detection** - The DIVA program intends to develop robust automated activity detection for a multi-camera streaming video environment.



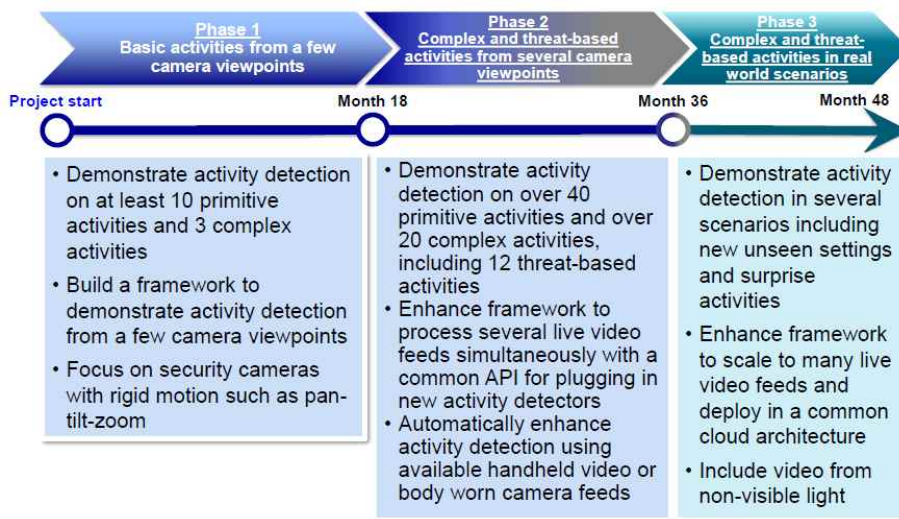
- (b) **Person/object detection** - DIVA activities will be enriched by person and object detection, as well as recognition at multiple levels of granularity.



출처: https://www.iarpa.gov/images/files/programs/diva/DIVA_PD_Briefing.pdf

DIVA는 2017년 7월부터 2021년 6월까지 총 48개월 동안 추진될 예정인데, 3단계로 구분되어 있다. 1단계에서는 사전에 정의된 단순 행위를 탐지하는 것부터 시작할 계획이다. 2단계에서는 탐지 가능한 행위의 종류를 확대하고, 3단계에서는 사전에 정의하지 않은 행위까지 탐지할 수 있는 수준까지 발전시킬 예정이다.

<그림 12> DARPA DIVA과제의 진행 단계별 연구 내용



출처: https://www.iarpa.gov/images/files/programs/diva/DIVA_PD_Briefing.pdf

측정 지표 기준의 DIVA의 단계별 목표는 아래의 표와 같다. 한편 동영상 스트림의 처리 속도는 1단계 15개월 동안은 실시간보다 5배 느린 것을 허용하다 2단계부터 실

시간 시스템으로 업그레이드 한다. 2단계 전반부 6개월 동안 2배 느린 것으로 향상시켰다가 후반 15개월 동안 실시간 처리를 완성한다. 이때 1단계와 2단계의 시험평가 시스템은 시장 구입가 기준으로 10,000 달러의 하드웨어를 이용한다. 3단계의 전반부와 후반에 각각 처리 속도를 2배속과 4배속으로 향상시키면서 동시에 하드웨어 구성도 7,000 달러와 5,000달러로 하락시킨다.

〈표 36〉 객체와 행위 탐지 측정 지표 기준 단계별 목표

구분	행위 탐지				객체 탐지			
	RFA 클래스A	RFA 클래스B	PMD 클래스A	PMD 클래스B	AFAR 장소1	APMD 장소1	AFAR 장소2	APMD 장소2
1단계	0.15	1.00	0.33	0.80	0.500	0.500	0.500	0.500
2단계	0.03	0.20	0.20	0.50	0.200	0.200	0.100	0.400
3단계	0.01	0.10	0.10	0.25	0.100	0.100	0.033	0.300

주1) RFA(Rate of False Alarms per Minute of Video), PMD(Probability of Missed Detection)

주2) 클래스A는 클래스B보다 탐지가 쉬운데 평가자 일치도, 기준 성능, DIVA 성능을 종합하여 결정

주3) AFAR(Average False Alarm Rate), APMD(Average Probability of Missed Detection)

출처: DIVA BAA

DIVA가 기존의 과제와 다음과 같은 점에서 차이가 난다. 첫째, 기존 과제들 대부분 (예: IARPA의 ALADDIN)이 유튜브(YouTube) 스타일의 동영상을 사후에 분석하는데 반해 DIVA는 CCTV로부터 입수한 동영상을 실시간으로 분석한다. 둘째, CCTV 동영상을 분석하기 때문에 인식 대상이 화면의 중앙이나 앞에 등장하지 않고 구석이나 뒤편에 등장할 수도 있는데, 이런 조건에서도 인식할 수 있어야 한다. 그리고 화면 해상도가 낮아도 인식할 수 있어야 한다. 셋째, 기존에는 단일 카메라 분석이었던데 반해 DIVA는 멀티 카메라를 대상으로 하고 있다. 여러 CCTV가 설치된 경우가 증가하는 점을 감안할 때 현실에 적합한 형태로 고도화되어 가고 있다고 볼 수 있다. 넷째, 객체 인식도 기존의 이미지넷(ImageNet)이나 마이크로소프트의 COCO¹⁹⁶⁾에서 다루지 않는 키워드로 확장할 것이다. 예를 들면, 무기(weapon), 총(guns), 지도(map) 등과 같이 이미지넷이나 COCO에 포함되지 않는 테러 관련해서 중요한 키워드를 추가할 것이다.

196) Common Objects in Context

2. 국내 관련 연구

DIVA와 관련된 국내 연구의 대표적인 연구로 현행 진행하고 있는 중장기 대형 과제 중 하나인 딥뷰가 있다. 딥뷰의 총괄과제는 응용 분야로 재난 재해의 조기 감지를 대상으로 한다. 이때 과제 제목에는 위험이 포함되어 있지 않지만 연구 목표에는 재난 재해와 함께 위험이 제시되어 있다. 따라서 딥뷰의 응용 분야가 DIVA의 응용 분야인 테러 감지와 일치한다고도 볼 수 있다. 실제 딥뷰의 사업책임자는 최근 언론과의 인터뷰에서 CCTV와 방송 분야를 응용 분야가 될 것이라고 밝혔다. 다만 딥뷰의 목표에 실시간 분석이 명시되어 있지 않다는 점에서 DIVA와 차이가 난다.¹⁹⁷⁾

딥뷰는 총괄과제와 4개의 세부과제로 구성되어 있는데, DIVA와 관련이 깊은 것은 2세부과제와 4세부과제이다. 2세부과제는 CCTV와 같이 다수의 카메라가 연동하는 상황에서 동일 객체를 전역적으로 추적하는 기술을 개발하고, 4세부과제는 싸우기, 껴안기 등 사람과 사람 또는 사람과 객체와의 상호작용을 탐지하는 것을 개발한다. 다만, 복수 카메라가 아니라 단일 카메라를 대상으로 한다는 것이 DIVA와 다르다. 딥뷰의 총괄과제와 4개 세부과제의 과제명과 수행기관, 과제책임자, 수행기간, 지원금액은 다음과 같다. 참고로 3세부과제는 위성 데이터의 분석 기술을 개발하는 과제인데, 2차년도까지 수행하다 조기에 종료되었다.

- (딥뷰-총괄) 대규모 실시간 영상 이해 기반의 시각지능 플랫폼 개발 및 재난 재해 조기 감지·예측 기술 개발 (ETRI 박경/박종열, ' 14.4.1~' 24.2.29, 1.0/1.0/1.0/1.0억원)
- (딥뷰1세부) 실시간 대규모 영상 데이터 이해·예측을 위한 고성능 비주얼 디스커버리 플랫폼 개발 (ETRI 박경/박종열, ' 14.4.1~' 24.2.29, 38.0/38.0/35.5/35.5억원)
- (딥뷰2세부) 대규모 실시간 영상 이해 기반의 시각 지능 플랫폼 개발 (광주과학기술원 전문구, ' 14.4.1~' 24.2.29, 20.0/20.0/18.7/18.7억원)
- (딥뷰-3세부) 환경 재난·재해의 조기 감지를 위한 다형 영상 및 복합 센서 데이터 기반의 분석·예측 기술 개발 (계명대학교 박상원, ' 14.4.1~' 16.2.29, 15.0/15.0억원)
- (딥뷰-4세부) 예지형 시각 지능 원천 기술 개발 (서울대학교 이경무 / POSTECH 김대진, ' 14.4.1~' 24.2.29, 3.0/6.0/6.0/6.0억원) *2차년도까지 경쟁으로 진행하다 통합

197) <http://www.boannews.com/media/view.asp?idx=55104>

최근에는 객체 추적과 관련하여 다중카메라, 심지어 카메라들 화면 사이의 중첩이 없는 환경에서의 보행자 동선 추적에 관한 기초연구까지 진행하고 있다. 동 과제는 2019년까지 3년 동안 지원되는데, 금액 연 5,000만원으로 소규모 연구이다. 한편 미디어로부터의 객체와 동작을 자동으로 인식하기 위한 단기 과제가 지원되기도 하였다. 각 과제의 과제명과 수행기관, 과제책임자, 수행기간, 지원금액은 다음과 같다.

- 비중첩 다중카메라 환경에서 딥러닝 기반 보행자 동선 추적 요소 기술 연구 (전북대 박동선, ' 16.11.1~' 19.10.31, 5,000만원/4,167만원)
- 온톨로지-기반 미디어 서술 및 미디어 마이닝을 위한 Deep-Learning을 이용한 객체검출 및 동작인식 기술 (연세대 최윤식, 2017.6.1.~2017.12.31., 5,000만원)

3. 향후 기획 방향

DARPA의 VIRAT은 2009년부터, IARPA의 ALADDIN은 2010년부터 시작되었다. 그리고 멀티 스트리밍 카메라 환경에서의 실시간 행위 탐지를 목적으로 하는 IARPA의 DIVA는 2017년 7월부터 시작되었다. 한편, 우리나라에서는 2014년부터 10년 장기 과제로 답류가 시작되었으며, 비중첩 다중카메라 환경에서의 보행자 동선 추적 등과 같은 도전적인 과제도 시작되었다. 그러나 DIVA가 복수 카메라 환경과 실시간 분석, 그리고 실질적인 현실 문제를 선정하고 정의한 것에 비추어볼 때 미국과 우리나라의 연구 목표에 격차가 존재한다. 지금 이 시점에서 이 차이를 그대로 방치하고 적절한 대응책을 마련하지 않는다면 기술 선도국과의 기술 격차는 앞으로 더 크게 벌어질 것이 자명하다. 따라서 지금 이 시점에 기술 선도국과의 대응한 경쟁을 위해 다중카메라 환경에서의 실시간 행위 탐지 연구과제의 기획을 적극적으로 검토해야 한다.

만약 우리나라에서 다중카메라 환경에서의 실시간 행위 탐지를 기획한다면 시각지능 분야, 특히 정보·감시·정찰 분야(ISR¹⁹⁸)에서 가장 도전적인 연구과제가 될 것인데, 다음의 같은 점을 고려하여 추진할 필요가 있다. 첫째, 응용 분야에 관한 것이다. 우선 카메라 환경을 명확하게 정의하고 응용 분야를 구체화하여 탐지에 필요한 행위를 사전에 정의해야 한다. 그런 측면에서 DIVA에서와 마찬가지로 정부 시설의 보호를 응용 분야로 선정하는 것을 고려해 볼 수 있다. 테러 행위의 유형이 상대적으로 많지 않고 테러를 사전에 탐지할 수 있다면 그 효과가 매우 크기 때문이다. 만약 정부 시설에 대

198) Intelligence, Surveillance and Reconnaissance

한 테러를 응용 분야로 선택한다면 DIVA에서 정의하고 있는 관심 행위 목록을 그대로 활용할 수 있다. 또 다른 예로는 자율주행차를 들 수 있다. 카메라를 통해 차량 주변의 행위나 상황 변화를 즉각적으로 탐지한 후에 다른 데이터, 예를 들면 라이더(LiDAR), 레이더(Radar) 등으로부터의 데이터를 종합하여 돌발 상황에 대응할 수도 있다. 그리고 지금은 자율주행차에 카메라가 1대 또는 전후방용으로 2대가 장착되는 형태로 개발되고 있지만 향후 카메라 가격이 하락하면 그 수는 많아질 수도 있다. 그렇게 되면 DIVA에서 정의한 다중카메라 실시간 데이터 분석의 조건에 부합된다. 한편 다른 응용 분야를 선정한다면 DIVA에서와 마찬가지로 데이터 분석에 앞서 관심 행동을 사전에 정의하는 것이 선행되어야 한다. 이때 탐지하고자 하는 행위 유형이 너무 다양하면 지금의 기술로는 개발에 한계가 있다.

둘째, 데이터 확보에 관한 것이다. 새롭게 기획되는 과제가 성공하기 위해서는 동영상 데이터셋의 확보가 필수조건이다. 그런데 최근 구글과 MIT-IBM Watson AI Lab에서 동영상 데이터를 공개하였는데 중장기 과제 초기에 공개 데이터를 적극적으로 이용하는 것을 고려해 볼 수 있다. 2016년 9월27일에 구글에서 8백만개의 동영상 데이터셋인 YouTube-8M Dataset¹⁹⁹⁾을 공개하였다.²⁰⁰⁾ 동 데이터 셋은 세계적으로 최소 1,000번 이상 시청한 동영상 클립이고, 동영상의 길이는 120초에서 500초 사이이다. 2017년 2월에는 동 데이터를 이용하는 것을 조건으로 하여 총상금 10만 달러의 공개 경연(Google Cloud & YouTube-8M Video Understanding Challenge)²⁰¹⁾을 발표하였고, 2017년 6월2일까지 접수된 팀들을 대상으로 경연이 진행되어 6월30일에 수상자들이 결정되었다. 2017년 12월에는 MIT-IBM Watson AI Lab에서 Moments in Time Dataset²⁰²⁾을 공개하였다. 동 데이터셋은 3초짜리 짧은 동영상 클립 약 100만 개로 구성되어 있다.²⁰³⁾ 그러나 공개 데이터를 이용할 수 있음에도 불구하고 궁극적으로 한국 사람들의 특징적인 행위를 탐지하기 위해서는 독자적인 데이터셋, 또는 공개 데이터셋을 보완할 데이터셋을 확보해야 한다.

DIVA는 독립적인 시험평가팀을 두고 있는데, 그 팀에서 평가에 필요한 데이터셋을 평가 최소 1개월 전에 수행기관에 제공하도록 되어 있다. 그런데 48개월의 3단계 진행 동안 평가는 수행기관의 자체 평가와 시험평가팀의 외부 평가로 구성된 5회의 평가가 이루어지게 된다. 수행기관은 자체 평가를 실시하고 시험평가팀에 평가결과를 통보한

199) <https://research.google.com/youtube8m/>

200) <https://research.googleblog.com/2016/09/announcing-youtube-8m-large-and-diverse.html>

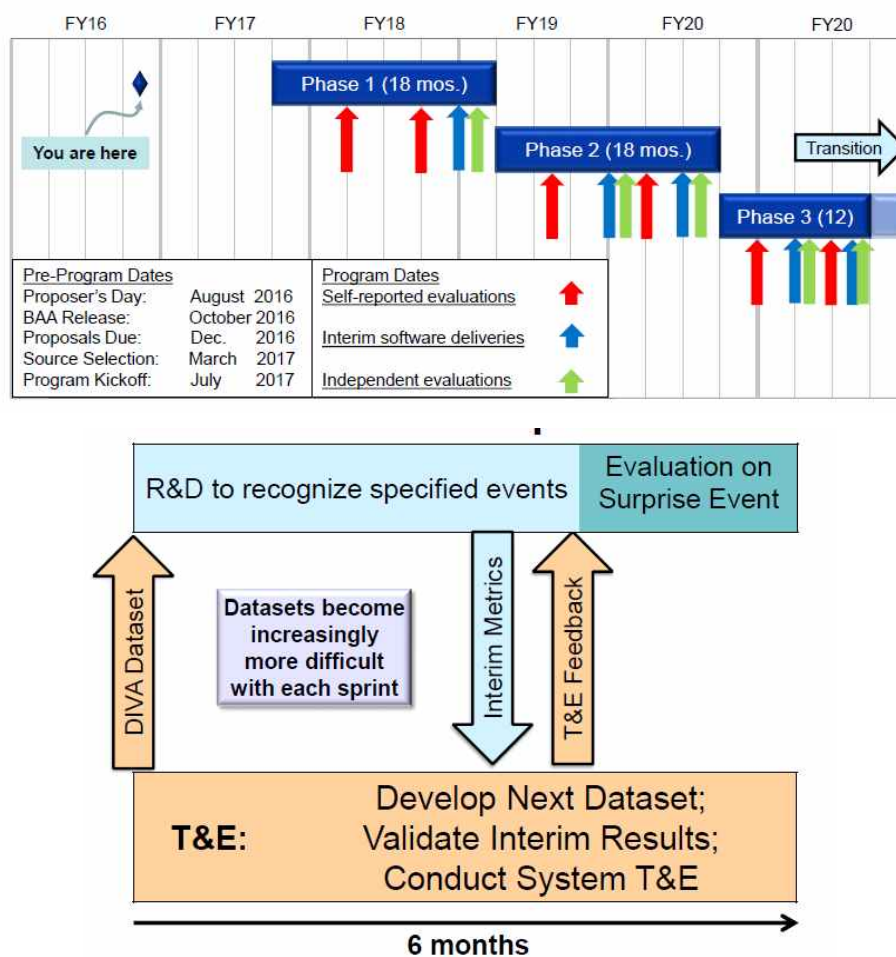
201) <https://www.kaggle.com/c/youtube8m>

202) <http://moments.csail.mit.edu/>

203) http://moments.csail.mit.edu/data/moments_paper.pdf

다. 그 다음에 수행기관이 개발한 소프트웨어를 시험평가팀에 제출하면 시험평가팀은 자체 개발한 평가 데이터를 이용하여 평가를 실시한다. 5회의 평가는 1단계의 1회, 2단계의 2회, 3단계의 2회이다. 이 외에 1단계에서 최초에 자체 평가 통보가 1회 추가된다. 결과적으로 1단계의 후반부부터 6개월 단위 스프린트를 실행하는 꼴이다. DIVA의 평가 일정과 6개월 단위의 스프린트 수행체계는 아래의 그림과 같다.

<그림 13> DIVA의 평가 일정 및 6개월 단위 스프린트 수행체계



출처: DIVA Proposers' Day 발표자료

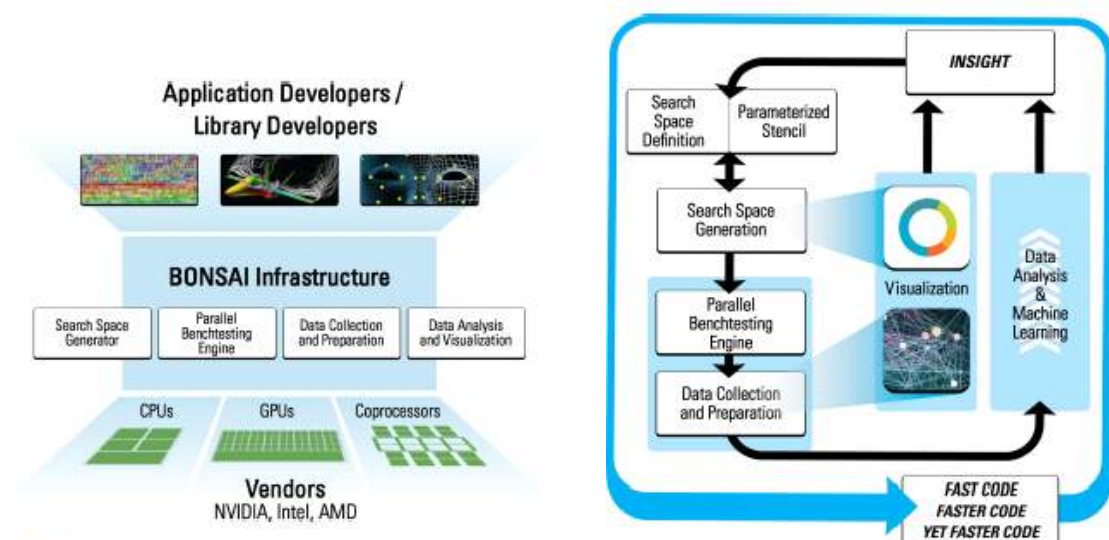
제3절 계산 커널의 병렬 자동 튜닝 오픈소스 소프트웨어

1. NSF BONSAI

최근 슈퍼컴퓨터들은 계산 속도를 향상시키고 비용을 절감하기 위해 매니코어 기반의 프로세서를 사용하고 있다. 그러나 매니코어 기반의 프로세서는 하드웨어 시스템의 가격을 낮추는데 성공하였으나 대신 소프트웨어를 작성하고 최적화하는 복잡성을 크게 증가시켰다. 또, GPU와 코프로세서의 하이브리드 시스템이 대세가 되면서 서로 다른 계산 커널을 최적화하는 것은 한층 어려워졌다. 수백만 가지의 커널 환경 조합 중 최적의 해를 찾는 일은 많은 시간이 소요된다.

병렬 자동 튜닝에 의한 소프트웨어 최적화가 해결책 중에 하나인데, 미국 NSF에서는 오래 전부터 이와 관련하여 여러 연구과제들을 지원해 왔다. 그 중 하나가 ‘BONSAI: An Open Software Infrastructure for Parallel Autotuning of Computational Kernels’ 이다. BONSAI의 최종 목표는 GPU 가속기와 매니코어 프로세서 기반의 시스템의 계산 커널(computation kernel)의 최적화 프로세스를 획기적으로 단축시키는 것이다. 이 연구과제는 검색 공간 생성기(Search Space Generator)와 병렬 벤치마킹 엔진(Parallel Benchmarking Engine), 데이터 수집 및 준비(Data Collection and Preparation), 데이터 분석 및 시각화(Data Analysis and Visualization)로 구성되어 있다.

<그림 14> NSF BONSAI의 구성과 프로세스



출처: https://figshare.com/articles/_BONSAI_an_Open_Software_Autotuning_Infrastructure/4668958

BONSAI는 벤더에서 제공하는 라이브러리와 오픈소스 라이브러리를 모두 활용한다. 예를 들면, 데이터 수집에는 엔비디아의 CUPTI API, AMD의 CodeXL, 인텔의 VTune, 오픈소스의 PAPI를, 데이터 분석에는 오픈소스 기반의 PyData stack, R, MLlib와 같은 Spark 툴을 활용할 것이다. 그리고 BONSAI의 연구결과는 오픈소스코드로 공개되어 대학과 기업에 모두 이용할 수 있게 할 예정인데, 대학 등에서 사용하는 클러스터부터 클라우드 컴퓨팅, 그리고 슈퍼컴퓨터까지 규모에 상관없이 모두 적용할 수 있게 하는 것을 목표로 하고 있다.

BONSAI는 2016년 11월1일부터 시작했고 2019년 10월31일에 종료될 예정이다. 총 지원 예산은 49만 9,977달러이다. 과제 책임자는 테네시대학의 Jakub Kurza(H-Index: 17)이다.²⁰⁴⁾²⁰⁵⁾ 한편, BONSAI는 2013년 8월1일부터 2016년 7월31일까지 기초연구로 진행된 Bench-testing Environment for Automated Software Tuning (BEAST)의 후속 연구이다. BEAST는 하이브리드 시스템에서의 계산 커널 성능을 탐색하고 최적화할 수 있는 프레임워크를 개발하는 것을 목표로 하였다. 이때, 완전 자동화가 아닌 반자동화를 지향하였다. BEAST의 과제책임자 역시 Jakub Kurza이고 지원 예산도 49만 9,977달러로 18달러밖에 차이가 나지 않는다.²⁰⁶⁾²⁰⁷⁾²⁰⁸⁾

2. 국내 관련 연구

국내에서는 계산 커널의 병렬 튜닝에 관한 연구는 드물다. 1억원 미만의 소규모 연구만이 진행되었다. 특히, 자동 튜닝에 관한 연구는 최근에서야 시작되었다. 2014년에 시작된 ‘매니코어 기반 초고성능 스케일러블 OS 기초 연구’에 자동 튜닝에 관한 연구가 일부분 포함되었고, 2016년에 시작된 ‘이기종 고성능 컴퓨터에서의 자동 성능 최적화 연구’가 첫 독립과제라고 할 수 있다. 관련된 국내 연구과제는 아래와 같다.

- 병렬성 극대화를 위한 소프트웨어 최적화 및 바이오 컴퓨팅 응용에 관한 연구 (연세대학교 정의영, '09.5.1~'12.4.30, 5,100/5,124/5,124만원, 교육과학기술부 일반연구지원사업)

204) https://www.nsf.gov/awardsearch/showAward?AWD_ID=1642441

205) <https://bitbucket.org/account/user/icl/projects/BON>

206) https://www.nsf.gov/awardsearch/showAward?AWD_ID=1320603

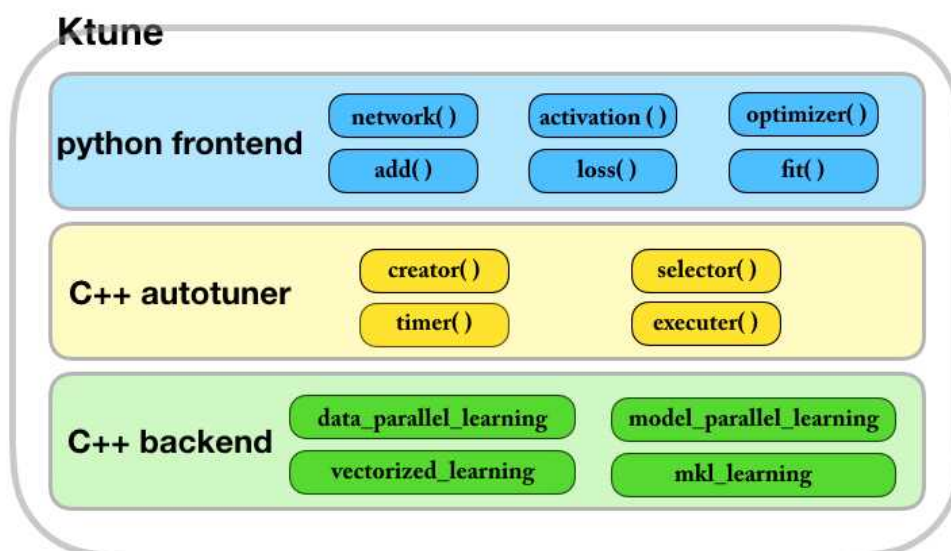
207) <https://bitbucket.org/account/user/icl/projects/BEAS>

208) http://www.netlib.org/utk/people/JackDongarra/WEB-PAGES/Batched-BLAS-2016/Day1/08_Piotr_Beastlang.pdf

- OpenACC를 활용한 고성능, 고 가용성의 이질적 병렬 프로그래밍 방법론 연구 (명지대학교 이명호, ' 12.9.1~' 15.8.31, 5,820/5,820/5,820만원, 교육부 이공학개인지초 연구지원사업)
- 내장형 시스템을 위한 병렬성 기반 실시간 멀티 코어 스케줄링 기술 개발 (한국과학기술원 신인식, ' 12.5.1~' 15.4.30, 4,924/4,885/4,924만원, 미래부 일반연구자지원 사업)
- (차세대OS기초연구센터) 매니코어 기반 초고성능 스케일러블 OS 기초 연구 (ETRI 정성인, ' 14.4.1.~' 22.2.28., 30.0/30.0/28.5/23.7억원)
- 이기종 고성능 컴퓨터에서의 자동 성능 최적화 연구 (명지대학교 이명호, ' 16.11.1~' 19.10.31, 5,000/4,167만원, 교육부 개인기초연구사업)

이 중 2014년부터 차세대OS기초연구센터 과제로 시작된 ‘매니코어 기반 초고성능 스케일러블 OS 기초 연구’의 연구 내용 중 일부에 오토튜너 연구가 포함되어 있다. 동 과제에서 개발한 오토튜너에는 ‘Ktune’라는 이름이 붙여졌는데 그 구조는 아래의 그림에서 보는 바와 같다.

〈그림 15〉 Ktune의 구성



출처: 매니코어 기반 초고성능 스케일러블 OS 기초 연구의 웹페이지

또한 2016년에 시작된 ‘이기종 고성능 컴퓨터에서의 자동 성능 최적화 연구’는 사용자 수준에서 적용 가능한 최적화 프레임워크 개발을 목표로 하고 있다. BONSAI의 전신이 BEAST의 연구 목표와 동일하다. 그런데 이 연구에서는 응용 프로그램을 제한하고 있다. 선정된 특정 응용 프로그램에 대해 최적화하는 것을 목표로 할 뿐 모든 응용 프로그램 대상으로 확대하는 것은 장기적인 주제로 삼고 있다. 즉, 선정된 응용 프로그램의 사용자들로 활용 그룹이 제한되어 있다.

3. 향후 기획 방향

NSF의 BONSAI과제의 전신인 BEAST가 2013년부터 시작되었고, 2015년에는 OpenCL²⁰⁹⁾ 커널의 자동 튜닝인 CLTune²¹⁰⁾²¹¹⁾²¹²⁾이 개발되어 공개되었다. 그런데 우리나라는 2016년이 되어서야 같은 연구 목표를 설정한 연구과제가 시작되었다. 그나마 예산도 매우 적고 개인연구의 성격을 띠고 있다. 연구과제의 차이까지 그대로 두면 몇 년 후 우리나라와 기술 선도국과의 기술 격차는 더 크게 벌어질 것으로 가능성이 크다. 따라서 지금 이 시점에 계산 커널의 병렬 자동 튜닝 오픈소스 소프트웨어 개발을 연구 내용으로 하는 연구과제의 기획을 적극적으로 검토할 필요가 있다.

더구나 오픈소스 기반의 오토 튜너는 사용자층을 넓히고 벤더들에 대한 종속 탈피 차원에서도 필요하다. 현재 특정 벤더들이 고가의 하드웨어와 함께 제공하는 애플리케이션 개발 환경을 이용할 수 있는 기업이나 연구자는 매우 제한적이고, 계산 커널의 최적화 툴 또한 벤더에 종속되면 마찬가지로 상황에 놓이게 된다. 따라서 다양한 벤더들의 제품을 수용할 수 있는 오픈소스 기반의 계산 커널 오토 튜너가 개발되어 공개되면 우리나라 사용자층 확대에 기여할 수 있을 것으로 기대된다. 그리고 궁극적으로는 현재 미국 업체들이 절대 다수를 차지하고 있는 하드웨어 벤더들에 대한 종속성도 약화시킬 수 있다.

만약 우리나라에서 계산 커널의 병렬 자동 튜닝 오픈소스 소프트웨어 과제를 기획한다면 고성능 컴퓨팅과 오픈소스 소프트웨어 분야를 모두 향상시키는 차세대 연구과제가 될 것인데, 다음의 같은 점을 고려하여 추진할 필요가 있다. 첫째, 연구 범위와 규모의 문제이다. NSF BONSAI는 3년 동안 약 50만 달러가 지원된다. 물론 BONSAI의 전

209) Open Computing Language

210) 네덜란드 SURFara HPCE 센터의 Cedric Nugteren와 Valeriu Codreanu 개발

211) <http://ieeexplore.ieee.org/document/7328205/>

212) <https://github.com/CNugteren/CLTune>

신인 BEAST까지 포함하면 연구과제 수행기간과 금액은 두 배로 늘어난다. BEAST와 BONSAI를 동시에 추진한다는 점에서 기간은 5년 이상이 되어 중장기 과제의 요건을 갖춘다고 볼 수 있다. 그러나 여전히 연 50억 원 내외를 투자해야 할 당위성을 확보하기 어렵다. 그런데 연구기간을 6년이 아니라 5년으로 압축하고, 2016년 시작한 명지대학교 이명호 교수의 연구에서 응용 프로그램을 제한하고 있는데 그 제한을 없애 응용 프로그램의 범위를 확대하면 연구과제의 규모 또한 확대될 수 있다. 그럼에도 불구하고 여전히 연 50억원 내외를 투자해야 하는지에 대한 검토는 필요하다. 만약 독립된 중장기 대형 연구과제로 추진하기 어렵다면 20억원 내외의 연구과제를 구성하여 현재 연 30억 원 정도가 지원되고 있는 ‘차세대OS기초연구센터’ 과제와 함께 추진하는 방법이 있다.

둘째, 오픈소스 커뮤니티와의 협력 또는 경쟁을 고려해야 한다. BONSAI는 기본적으로 오픈소스소프트웨어 프로젝트이다. 공개 소프트웨어를 이용하는 것뿐만 아니라 연구 결과도 오픈소스로 공개한다. BONSAI뿐만 아니라 CLTune 또한 마찬가지로 깃허브(GitHub)에 자신들이 연구한 것들을 공개하고 있다. 거의 모든 계산 커널의 오토 튜너 연구자들은 연구결과를 공개하고 있다. 따라서 우리나라가 처음부터 끝까지 개발하는 것 외에도 이미 개발된 것들을 확대하는데 기여하는 방안도 있다. 다만 연구과제의 본격적인 기획단계에서는 BONSAI나 CLTune 등의 공개 정도가 어느 정도인지 판단해야 한다. 만약 핵심적인 소스코드를 공개하지 않아 연구개발의 성공을 장담할 수 없다면 이미 개발된 오픈소스를 활용하는 것보다 직접 개발하는 것을 전략적으로 선택할 수 있기 때문이다.

셋째, 계산 커널의 오토 튜너에 대한 실질적인 수요를 발굴하고 선택해야 한다. 물론 대학에서 사용하고 있는 클러스터나 상용 클라우드, KISTI와 기상청에서 보유하고 있는 공공 슈퍼컴퓨터, 그리고 향후 자율주행차 등과 같은 임베디드 시스템에 이종 프로세서 기반 컴퓨팅 시스템의 탑재가 늘어날 것이다. 그리고 계산 커널의 병렬 자동 튜닝에 대한 수요는 증가할 것이다. 그런데 연구과제 기획에서 불명확한 대상의 응용보다는 명확한 것을 목표로 설정하는 것이 바람직하다. 그런 의미에서 기상청의 하이브리드 시스템으로의 전환은 좋은 기회가 될 수 있다. 기상청은 2011년부터 2019년까지 9년간 한국형 수치예보모델을 개발을 위해 (재)한국형수치예보모델개발단²¹³⁾을 출범시켜 사업을 추진해 오고 있다. 예비타당성조사에서 946억 8,400만원의 사업비가 책정된 대형 사업이다. 한편 2016년에 인텔 CPU 기반의 4호기를 도입하였는데, 2021년에 도

213) <https://www.kiaps.org>

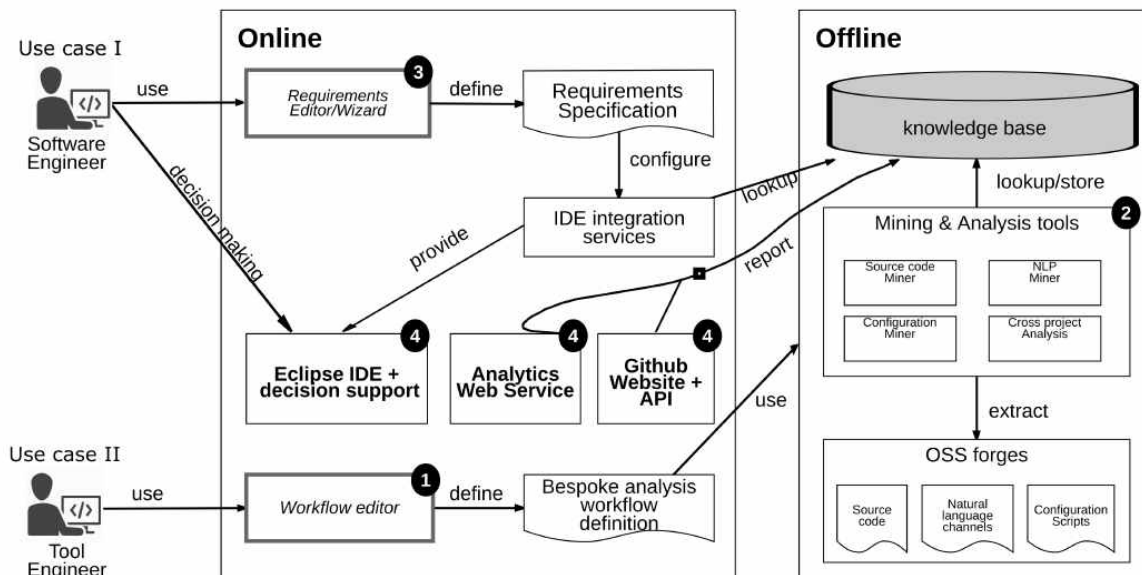
입 예정인 차기 시스템은 비용 절감 등 여러 이유로 GPU 가속기가 추가된 하이브리드 시스템을 도입할 가능성이 높다. 이때 현재 CPU 환경에 최적화된 한국형 수치예보 모델을 이중 프로세서에서 최적화시키는 문제가 발생한다. 따라서 기상청의 시스템 전환을 계산 커널 오토 튜너의 응용 목표로 적극적으로 고려해 볼 수 있다.

제4절 오픈소스 소프트웨어 커뮤니티 활동 및 코드 자동 학습

1. H2020 CROSSMINER와 DARPA MUSE

유럽연합은 2000년 12월에 첫 오픈소스소프트웨어(OSS) 전략을 수립한 이래 오픈소스소프트웨어의 활성화를 위해 지속적으로 노력하고 있고 H2020 연구개발과제로도 지원하고 있다. 그 중 하나가 2017년 1월1일부터 2019년 12월31일까지 3년 동안 CROSSMINER²¹⁴⁾²¹⁵⁾이다. 개발자들은 필연적으로 알맞은 오픈소스 컴포넌트를 선택해야 하는 문제에 직면할 수밖에 없다. 무수히 많은 컴포넌트 중 검토할 대상을 선택하여 비교해야 한다. 또 자신의 개발 환경과 요구조건에 맞게 변경해야 하는데, 이 모든 작업은 상당한 시간이 소요된다. 그래서 CROSSMINER는 개발자를 위한 오픈소스 커뮤니티 이용 지원 플랫폼을 지향하고 있다. 개발자들이 일일이 찾아 확인해야 하는 정보를 플랫폼에서 자동으로 분석하여 제공하는 것뿐만 아니라 개발자에게 가장 적합한 컴포넌트를 추천하는 것과 개발 환경에 통합하는 것까지 포함하고 있다.

〈그림 16〉 CROSSMINER 접근법



출처: Alessandra Bagnato 외 (2017) Developer-Centric Knowledge Mining from Large Open-Source Software Repositories (CROSSMINER)

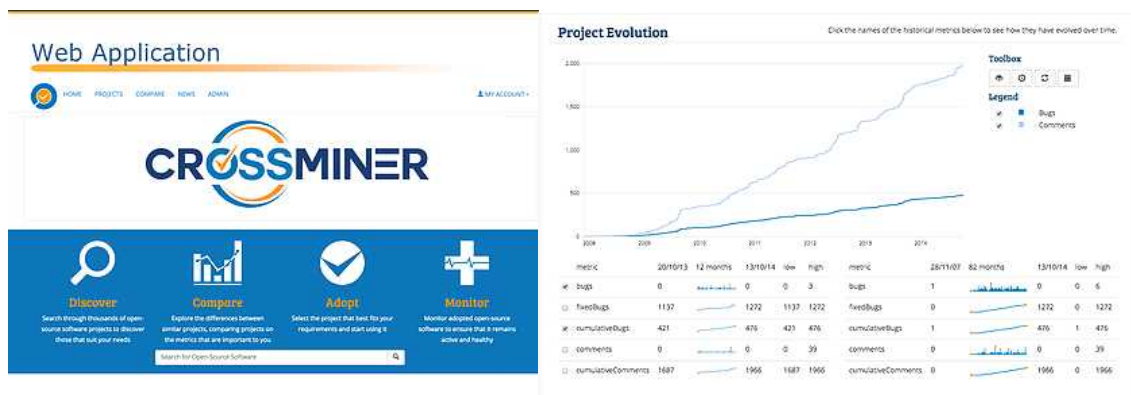
214) <https://www.crossminer.org/>

215) <https://www.dodwatchhub.eu/service-offer-catalogue/crossminer-developer-centric-knowledge-mining-large-open-source-software>

CROSSMINER의 기술적 구성 요소는 다음과 같은 여섯 가지이다. 그 중 네 가지는 앞의 그림 중 오프라인(offline)의 중간에 위치해 있는 마이닝 및 분석 도구들(mining and analysis tools)이다. 구체적으로는 소스코드 마이너(source code miner)와 자연어 처리 마이너(NLP miner), 시스템 환경 마이너(configuration miner), 프로젝트 관계 분석 툴(cross-project analysis tool)이다. 그 외에 워크플로우 기반 지식 추출기(workflow-based knowledge extractor)와 고급화된 통합 개발 환경(advanced integrated development environment)이 있다.

CROSSMINER엔 총 12개 기관이 참여하고 있다. 코디네이터(coordinator)인 영국의 글로벌 컨소시엄인 X/Open²¹⁶⁾을 비롯해서 영국의 요크대학교와 엣지힐대학(Edge Hill University), 프랑스의 Softeam과 OW2 컨소시엄 연합(Consortium Association), 독일에 위치해 있는 이클립스 재단 유럽, 그리고 이탈리아, 네덜란드, 스페인, 포르투갈, 그리스, 헝가리에서 각각 1개 기관이 참여하고 있다. 총 예산은 54,519,007 유로인데, 유럽 연합에서 전액 지원한다.

<그림 17> CROSSMINER 플랫폼 초기 화면 및 세부 화면



CROSSMINER는 2012년부터 10월1일부터 2015년 3월31일까지 진행된 OSSMETER²¹⁷⁾²¹⁸⁾의 후속과제이다. OSSMETER는 오픈소스소프트웨어 저장소(forge) 개발을 목표로 하지 않았다. 대신 이미 존재하는 저장소나 단체를 분석하는 메타플랫폼을 지향한다. 즉, SourceForge, Google Code, GitHub, Eclipse, Mozilla, Apache 등에 저장된 소스코드와 프로젝트 메타데이터, 커뮤니티 내 의사소통 데이터, 그리고 버그 보고 및 추적 시스템을 분석한다. OSSMETER의 총예산은 3,589.297유로인데, 그 중 2,603,000 유로는 유럽연합이 지원하였다.

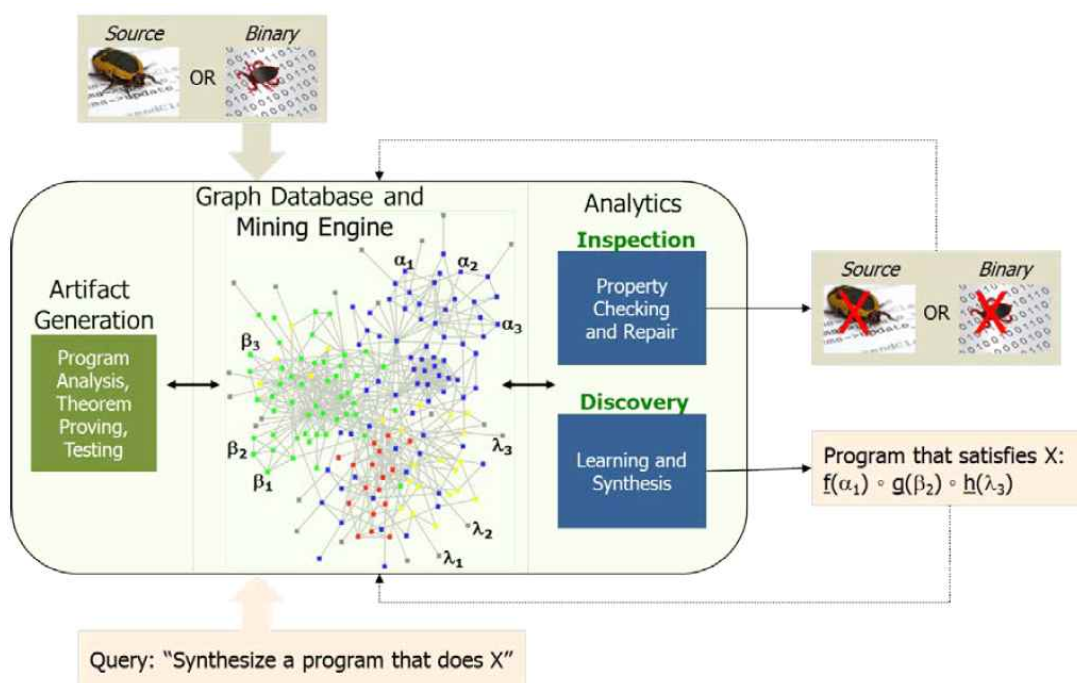
216) <http://www.opengroup.org>

217) Automated Measurement and Analysis of Open Source Software

218) http://cordis.europa.eu/project/rcn/105434_en.html

한편, DARPA에서는 오픈소스 컴포넌트 추천을 넘어 소프트웨어 코드를 학습하여 소프트웨어 품질을 개선하려는 시도를 하고 있다. 대표적인 것이 MUSE²¹⁹⁾²²⁰⁾과제이다. 지금까지 개발자의 경험이나 특화된 툴을 이용해서 소프트웨어 코드 품질을 제고하던 것을 기계 학습을 통해 자동화하려는 움직임이다. 기계 학습에는 이미 공개되어 있는 대용량의 오픈소스 코드를 이용하는데, 이 과제가 추구하는 바가 실현되면 중국에는 프로그램이 스스로 프로그램을 작성하고 디버깅을 될 수도 있다. 아래 그림에서 보는 바와 같이, MUSE 시스템은 버그가 포함된 소스코드 또는 바이너리 코드를 그래프 기반의 데이터베이스로 표현하고 학습에 필요한 요소를 추출할 수 있는 엔진을 탑재하고 있다. 규칙을 발견하는 분석 툴이 탑재되어 있는 것과 동시에 발견된 규칙을 이용하여 검사하여 디버깅을 한다.

<그림 18> MUSE 시스템 개념도



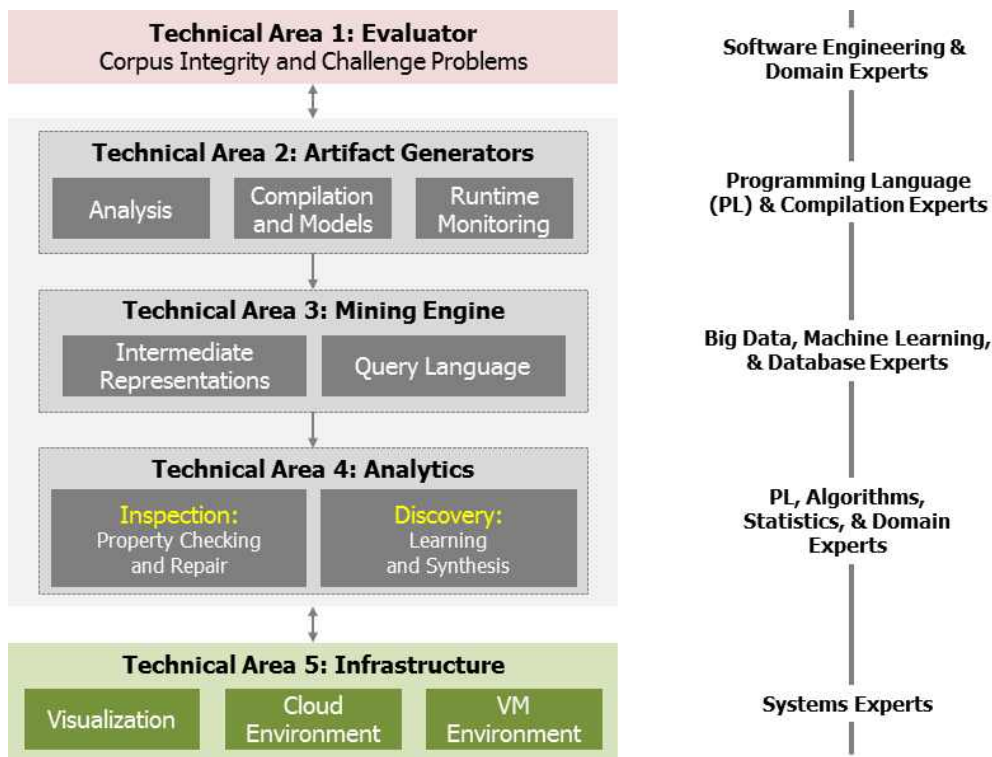
MUSE의 사업기간은 2014년 8월에 시작하여 2018년 11월말에 종료될 예정으로 총 52개월이다. 예산은 2014년의 450만 달러를 시작으로 2015년의 800만 달러와 2016년의 1,206만 9천 달러, 그리고 2017년과 2018년은 각각 1,300만 달러가 배정되어, 총 5,056만 9천 달러이다. MUSE의 세부 과제 구성은 아래의 그림과 같은데, 최소 9개의

219) Mining and Understanding Software Enclaves

220) <http://materials.dagstuhl.de/files/15/15472/15472.SureshJagannathan1.Slides.pdf>

기관이 이 사업에 참여하고 있다. 라이스대학과 Kestrel Institute가 소프트웨어 분석, 분석 툴, 프로그램 통합, 데이터베이스 구축 등을 주도하고 있으며, 조지아공대와 MIT CSAIL²²¹⁾, SRI, 콜로라도대학, 텍사스주립대 오스틴, 위스콘신-메디슨대학교, 그리고 소프트웨어 코푸스(corpus)를 담당하는 Leidos²²²⁾가 참여기관이다.

〈그림 19〉 DARPA MUSE 사업의 세부 과제 구성



출처: <https://www.darpa.mil/attachments/MUSEProgramOverview.pptx>

MUSE사업의 일환으로 Rice University는 PLINY 프로젝트(1,100만 달러, 4년)를 수행하고 있다. 이 프로젝트에는 MUSE의 참여기관이기도 한 텍사스 주립대 오스틴, 위스콘신 주립대 매디슨, 그리고 GrammaTech²²³⁾가 함께 참여하고 있다.²²⁴⁾

한편, MUSE는 연구결과를 오픈소스소프트웨어로 공개하는 것을 기본으로 하고 있다. 현재 일부는 오픈소스 라이선스로 공개되어 있다.²²⁵⁾ 대표적으로 MIT CSAIL에서 개발한 분석툴인 Prophet과 Sketch의 소스코드가 공개되어 있다.²²⁶⁾ 또 SRI International에

221) Computer Science and Artificial Intelligence Laboratory

222) 미국의 방산업체인 Science Applications International Corporation(SAIC)의 자회사로 뉴욕증시에 상장되어 있음

223) <https://www.grammatech.com/>

224) <http://pliny.rice.edu/>

225) <https://opencatalog.darpa.mil/MUSE.html>

226) <http://groups.csail.mit.edu/pac/patchgen/>

서 자바 프로그래머를 위한 분석틀인 PASCALI의 소소코드도 공개되어 있다.

오픈소스 소프트웨어 리파지토리로부터 데이터를 학습하려는 시도는 H2020의 CROSSMINER나 DARPA의 MUSE 외에도 유럽과 미국에서 활발하게 이루어지고 있다. 이와 관련된 연구들은 2010년을 전후에 빠르게 증가했다.²²⁷⁾²²⁸⁾ 2013년도 ICSE(International Conference on Software Engineering)에서 학습을 통한 자동 패치 생성(automatic patch generation)에 관한 첫 세션이 개최되었다.²²⁹⁾ 2015년부터는 ‘빅 코드(big code)’라는 용어가 사용되기 시작했다. 다른 키워드로는 ‘인텔리전트 소프트웨어 엔지니어링(intelligent software engineering)’²³⁰⁾이 있다. 그리고 2017년 10월에 처음으로 국제 워크샵이 개최되었다. 제1회 인텔리전트 소프트웨어 엔지니어링 국제 워크샵(WISE: International Workshop on Intelligent Software Engineering)이 제32회 IEEE/ACM의 ASE²³¹⁾의 일환으로 개최되었다. 여기에서는 미국의 마이크로소프트, 구글, 페이스북 뿐만 아니라 중국의 화웨이에서도 발표를 했다. 마이크로소프트의 Miltos Allamanis²³²⁾는 Machine Learning for Big Code and Naturalness를 이끌고 있다.²³³⁾

유럽에서도 비슷한 노력이 이어지고 있다. 유럽연합연구실(ERC)인 취리히공대의 소프트웨어 신뢰성 연구실, 이스라엘 테크니온대학의 소프트웨어 연구실, 영국 에딘버러 대학의 Machine Learning for the Analysis of Source code Text (MAST)²³⁴⁾ 프로젝트 등이 참여하고 있는 빅코드(Big Code) 프로젝트가 있다.²³⁵⁾²³⁶⁾ 취리히공대의 저자들은 빅코드에 관한 논문을 발표했다.²³⁷⁾

2. 국내 관련 연구

H2020의 CROSSMINER와 DARPA의 MUSE 등과 같이 미국이나 유럽에서 오픈소스 소프트웨어 활동이나 코드를 학습하는 움직임은 매우 활발한데 반해 우리나라에서는 관련 연구가 전무한 실정이다. 유일하게 최근 중앙대학교 이찬근 교수가 관련 연구를 진

227) <http://program-repair.org/bibliography.html>

228) <https://ml4code.github.io/papers.html>

229) <https://arxiv.org/abs/1408.2103>

230) <https://www.slideshare.net/taoxiease/intelligent-software-engineering-synergy-between-ai-and-software-engineering>

231) International Conference on Automated Software Engineering

232) <https://miltos.allamanis.com/>

233) <https://ml4code.github.io/>

234) <https://mast-group.github.io/>

235) <https://github.com/learnbigcode/learnbigcode.github.io/blob/master/community/people/index.md>

236) <http://learnbigcode.github.io/community/people/>

237) Predicting program properties from "Big Code"

행하고 있다. 대표적으로 2014년부터 시작된 데이터 마이닝 과제가 있다. 또, 2013년부터 2016년까지 진행한 정보통신기술인력양성사업 과제의 세부 연구 내용에 코딩 오류 탐지용 오픈 소스 코드 체커 개발 등이 포함되어 있다. 그러나 이 과제만이 규모가 있는 과제였고, 다른 과제들은 모두 1억원 이하의 개인연구들이다. 관련된 국내 연구과제는 아래와 같다.

- 딥러닝 기반의 지능형 버그 관리 도구 개발 (중앙대 이찬근²³⁸), 2017.11.1.~2020.10.31., 3,300만원, 미래부 개인기초연구사업 전략과제)
- 버그 담당자 배정 자동화를 위한 컨볼루션 신경망 기반의 선별 기법 (중앙대 이찬근, 2017.6.1.~2018.5.31., 5,000만원, 교육부 개인기초연구사업 기본연구)
- 소프트웨어 아키텍처 모듈-뷰의 복원과 분석을 위한 데이터마이닝 기반의 워크벤치 (중앙대 이찬근, 2014.5.1.~2017.4.30., 0.99/0.99/0.99억원, 미래부 개인연구지원사업 핵심개인연구)
- 고품질 융합 소프트웨어 개발 지원 도구 연구 (중앙대 이찬근, 2013.9.1.~2016.12.31., 6.40/8.35/8.00/7.20억원, 미래부 정보통신기술인력양성사업 대학ICT연구센터육성지원)
- 소프트웨어 리파지토리 마이닝에 기반하여 변경 클론을 이용한 포괄적인 변경 결합도 예측 기법 (2012.5.1.~2015.4.30., 경북대 이은주, 5,070/5,070/5,070만원, 교육과학기술부 여성과학자지원사업)

3. 향후 기획 방향

오픈소스소프트웨어의 중요성이 한층 고조되고 있음에도 불구하고 우리나라에서는 오픈소스코드를 활용하는데 한계가 있다. 미국과 유럽에서는 오픈소스 코드를 데이터 마이닝과 학습의 대상으로 인식하기 시작하는데 반해 우리나라에서는 그렇지 못하고 있다. 이런 인식의 차이가 연구과제의 차이로도 드러난다고 볼 수 있다. 그리고 미래를 위한 투자의 차이가 몇 년 후 우리나라와 기술 선도국과의 기술 격차는 더 크게 벌릴 수 있다. 따라서 지금 이 시점에 오픈소스 활동을 마이닝하거나 소스코드를 학습하는 연구과제의 기획을 적극적으로 검토해야 한다.

²³⁸) <https://sites.google.com/site/rtselab/real-time-software-engineering>

만약 우리나라에서 오픈소스 소프트웨어 커뮤니티 활동 및 코드 학습에 관한 연구과제를 기획한다면 오픈소스소프트웨어의 활성화에 기여할 것으로 기대되는데, 다음의 같은 점을 고려하여 추진할 필요가 있다. 첫째, 연구 범위의 문제이다. 비록 본 연구에서는 H2020의 CROSSMINER를 최종 후보로 선정하고 추진방안을 제시하고 있지만, 미국 DARPA의 MUSE까지 모두 포괄하는 형태로 확대할 필요가 있다. 오픈소스 커뮤니티에 대한 데이터 마이닝만으로는 중장기 대형 과제가 되는데 한계가 있는데 소스코드 학습까지 포함하면 그 한계를 벗어날 수 있다. 또 서로 연관 있는 두 분야를 동시에 진행함으로써 기술 선도국과의 격차를 획기적으로 줄이는데도 기여할 것이다. 이때 소프트웨어 엔지니어링 연구자들과 저작권위원회처럼 공개소프트웨어 데이터베이스를 관리하는 기관과 협력이 이루어지는 것이 바람직하다.

둘째, 실제 개발자들이 이용할 수 있는 인프라로 구축하는 것이다. 최근 정부 재원이 투자되는 소프트웨어 연구개발과제들의 결과를 공개하려고 하는 움직임이 있는데, 전제 조건은 라이선스나 품질이 제대로 관리되는 것이다. 그런데 지금은 소스 코드의 품질 관리는 개발자나 개발자가 속한 기업의 몫이다. 물론 정부에서 제공하는 서비스도 있다. 예를 들면, 저작권위원회에서는 라이선스 검증 서비스(CodeEye)를 제공하고 있는데, 공개소프트웨어 라이선스 데이터베이스를 구축해서 지속적으로 업데이트하고 있다. 또 공공기관이나 중소기업의 공개소프트웨어 도입을 돕는 컨설팅 서비스도 제공하고 있다. 그러나 이들 서비스는 자동화되어 있지 않아 비용 대비 효과가 제한적이다. 따라서 연구개발과제를 통해 자동화된 소스 코드 분석 툴이 개발되면 저작권위원회 등의 서비스로 제공될 수 있을 것이다. 한편 이런 요구사항은 대규모 온라인 공개강좌(MOOC)의 소프트웨어 프로그래밍 과목에서 수강생들의 프로그래밍 결과물 평가에도 동일하게 적용할 수 있다.

마지막으로 이미 활발하게 진행하고 있는 미국과 유럽의 전문가들과의 교류와 협력을 강화하는 것이 바람직하다. 특히 우리가 벤치마킹하고자 하는 미국과 유럽의 연구개발과제들이 대체로 연구 결과를 적극적으로 공개하고 있기 때문에 후발 주자인 우리나라에서 그들과의 협력으로부터 얻을 것이 많다. 특히 급부상하는 연구자들과 적극적으로 협력할 필요가 있다. 예를 들면, 현재 MIT의 박사과정 학생인 Long Fan(H-Index: 11)²³⁹⁾을 들 수 있다.

239) <https://pdfs.semanticscholar.org/53c0/61f4ec5ca140c7297e54361608533dce409.pdf>

제7장 결론

본 연구에서는 제4차 산업혁명에 대응하기 위해 선도국과의 SW기술 격차를 해소하거나 역전시키기 위해 필요한 R&D 추진체계와 중장기 대형 연구개발과제 후보를 선정하여 제시하였다.

빠른 시간 내에 SW 선진국으로 도약하기 위해서는 DARPA 사례분석에서 살펴봤듯이 고위험-고성과를 지향하는 새로운 R&D 추진체계가 필요하다. 이를 위해 실질적 수요에 기반한 명확한 목표 설정 및 과제기획이 이루어져야 한다. 또한 과제책임자가 강한 리더십을 발휘하여 불확실성 유연하게 대처하고 성과를 창출할 수 있도록 과제 운영에 대한 권한과 재량권을 대폭 강화해야 한다.

과제 후보의 선발을 전형적인 상향식 수요조사나 키워드 중심의 분야 선정을 따르지 않고 기술 선도국에서 최근 진행하고 있는 사업 또는 과제를 중심으로 연구팀에서 선발하여 외부 전문가들에게 선택을 의뢰하는 방식을 택하였다. 해외 연구개발과제를 벤치마킹하는 것에서 출발하다 보니 추진 방법에 있어서도 몇 가지 특징을 보인다.

첫째, 자연스럽게 도전적인 연구 목표를 설정하게 된다. 기술 선도국에서 설정한 목표 중에서도 기술 격차 해소 차원에서 의미 있는 후보를 선택하기 때문에 기술적으로 가장 최첨단의 목표를 확인할 수 있다. 그 이후 기존 우리나라의 연구개발과제와 비교하면서 지금보다는 더 도전적인 연구 목표를 설정하게 된다. 이것은 본 연구에서 당초 의도했던 대로 기술발전의 궤적을 고려하여 연구 목표를 설정하고 추진방향을 기획할 수 있다는 것을 의미한다. 즉, 기술 선도국과의 기술 격차가 존재하는 상황에서 기술 선도국에서 최첨단의 추진 방향, 그것도 구체적인 기술 목표를 설정한 연구과제들을 종합적으로 검토함으로써 기술 진화의 방향 또는 다양성을 충분히 검토할 수 있다.

둘째, 현장 적용에 초점을 두게 된다. 미국의 DARPA와 IARPA 과제에서 현장 적용을 필수 요건으로 가지고 있는데, 본 연구에서 후보 검토 단계부터 그 부분을 충분히 파악하기 때문에 후보 선정 후 기획과정에서 그 내용을 충분히 반영할 수 있다. 무엇보다 참고하는 사업이나 과제들이 현장, 그것도 공공 수요에 기초해서 기획한 것이어서 그 응용 사례가 매우 구체적인 경우가 많다.

셋째, 기존에 우리나라에서 진행되고 있는 관련 연구과제들과의 협력과 차별화를 강조하게 된다. 후보를 검토하는 과정에서 국내 연구과제의 현황을 체계적으로 파악하기 때문이다. 연구 목표와 내용, 수행방법까지 모든 것이 똑같은 연구개발과제는 없더라

도 어느 정도 연관된 주제를 취급하는 연구개발과제는 존재할 수밖에 없다. 그런데 지금까지는 마치 백지에서 구상하는 것처럼 과제를 기획한 후 중복성 검토를 추후에 거치게 된다. 그에 반해 본 연구에서는 기존 연구과제들을 사전에 파악한 후 연계나 차별화를 고민하여 기획 방향을 제시할 수 있었다.

마지막으로 무엇보다 해당 분야의 연구 그룹, 그 중에서도 선두 그룹 그리고 공공 재원을 이용하여 연구를 진행하는 연구 그룹을 사전에 파악함으로써 기획 단계부터 국제 협력을 적극적으로 검토할 수 있다. 예를 들면, NSF나 H2020의 과제와 같이 연구결과를 적극적으로 공개하는 경우에는 직접 협력도 가능할 것으로 기대한다. 비록 DARPA나 IARPA의 경우 연구결과가 공개되지 않더라도 연구수행 주체를 확인할 수 있어 그들의 연구 현황이나 방향을 전문가의 입장에서 판단할 수 있다.

이렇듯 기술 선도국의 사업이나 과제를 대상으로 후보를 발굴하고 기획하게 되면 보다 기획 초기부터 구체적인 정보를 바탕으로 기술적 도전 목표의 수립이나 응용 사례의 구체화, 국내 연구과제 진행 상황을 고려한 연구 그룹의 형성, 그리고 해외 전문가를 사전에 파악함으로써 국제 협력 전략 수립이 가능하다.

위와 같은 장점을 바탕으로 본 연구에서 제시한 예비 후보 선발과 후보 검토 프로세스는 정부의 소프트웨어 과제 기획 프로세스에 접목할 수 있을 것으로 기대된다. 일반적으로 수요조사가 2사분기에 이루어지는데, 그 전에 기획팀에서 기술 선도국의 사업이나 과제를 대상으로 예비 후보를 선발하고, 나아가 후보 선정까지 이루어진다면 수요조사에서 제안된 것들과 함께 검토하여 과제 기획의 품질을 제고할 수 있을 것이다. 특히, 정부의 연구개발관리 전담기관은 충분한 전문가 풀을 확보하고 있어 후보 검토 과정을 보다 잘 이끌 수 있을 것으로 기대된다.

한편, 본 연구에서 제시한 예비 후보 선발과 후보 검토 프로세스는 최종 후보 선정이라는 본래의 목적 외에 부가적으로 활용될 가능성도 가지고 있다. 예를 들면 다음과 같다. 첫째, 해외 연구과제 동향 정보 수집에도 활용할 수 있다. 전문가들에게 후보와 최종 후보 검토를 의뢰하기 위해서 해외에서 추진하고 주요 연구과제를 1페이지의 설명 자료로 요약하였는데, 이것을 체계적으로 진행하면 동향 정보를 수집하는 효과를 얻을 수 있다. 본 연구에서는 총 108개의 예비 후보가 선발되었고, 각각 1페이지로 설명자료를 작성하였다. 즉, 108개의 연구개발과제 정보가 관리된 것이다. 물론 예비 후보 선정에 활용한 3,332개의 목록도 함께 관리하였다. 또, 후보 선발과 후보 선정을 전문가에게 의뢰할 때까지는 해외 과제에 대한 설명만 담았는데, 최종 후보 검토에서는 국내 연구개발과제에 대한 정보도 함께 추가하였다. 이렇듯 해외 동향을 기획의 단계

가 심화될 때마다 추가적인 정보를 더함으로써 제한된 기획 자원을 합리적으로 활용할 수 있는 장점도 가지고 있다.

둘째, 기술 수준 조사로의 대안이 될 수 있다. 현재 기술 수준 조사는 전문가의 설문 을 중심으로 이루어지고 있다. IITP는 매해, KISTEP은 2년마다 실시하고 있다. 비록 이들 기관의 기술 수준 조사가 정량화하는데 공을 들여 발전시켜왔지만 여전히 논문이나 특허 분석 등과 같이 계량적인 접근법을 보다 적극적으로 활용해야 한다는 목소리가 있다. 많은 연구자들이 이와 관련된 연구도 진행하고 있고, 비록 전문가 설문과 통합하지는 않았지만 IITP와 KISTEP에서도 별도의 계량적 분석도 함께 실시하고 있다. 그런데 본 연구를 진행한 경험에서 새로운 제안을 하고자 한다. 정부에서 기술 선도국에서 기획한 연구개발 사업이나 과제와 유사한 과제를 기획하여 그 연구 목표나 내용, 그리고 진행 상황, 또 최종 결과 등을 비교하여 직접 비교하는 방식이 있다. 직접 비교하는 방식이어서 의미가 있을 것으로 기대된다. 이때 동일한 연구 목표나 내용을 고집할 필요는 없을 것이다. 다소 차이가 있더라도 그 차이를 비교에 충분히 반영할 수 있다면 기술 수준 조사로서 가치가 있을 것 같다.

셋째, 사회문제해결이나 공공혁신의 주제 선정에도 활용할 수 있다. 최근 우리나라에서는 정부의 연구개발 투자가 산업 육성뿐만 아니라 사회문제 해결에 보다 적극적으로 나서야 한다는 목소리가 높다. 그런데 사회문제해결의 연구 주제 대부분이 기존의 융합 연구와 상당히 겹치고 사회문제의 제시를 누가 해야 하는가 등에 대한 논란이 제기 되고 있다. 즉, 사회문제해결형 연구개발은 지금까지와는 다른 기획 프로세스가 필요할 것이라는 것을 암시한다. 그런데 본 연구에서는 기술 선도국들이 구체적인 응용을 전제로 한 연구개발과제들을 충분히 검토한 바 있다. 그리고 그들이 사회문제해결의 연구개발과제를 어떻게 기획하고 있는지를 가늠해 볼 수 있다. 특히, 유럽연합의 경우에는 H2020의 최상위 목표 중 하나가 사회 도전과제의 해결이고, 7개 분야를 선정하여 기술적 해결 대안을 찾고 있다. 또 DARPA나 IARPA는 공공 수요를 PM이 주도적으로 찾고 확인하는 과정과 연구자 커뮤니티로부터 기술적 가능성에 대한 의견을 수렴하는 과정을 거친다. 따라서 기술 선도국이 사회문제해결을 위해 어떤 문제에 정부 재원을 투자하고 있고, 그 연구개발과제를 어떻게 기획하는 지 검토하는 것은 매우 의미 있을 것이다.

참 고 문 헌

국내 문헌

- 정보통신기술진흥센터(IITP) (2016.2) 2015년도 ICT 기술수준조사 보고서
정보통신기술진흥센터(IITP) (2017.9) 2016년도 ICT 기술수준 조사보고서
이근 (2014.12) 경제추격론의 재창조, 오래
송위진 외 (2006) 탈추격형 기술혁신체제의 모색, 과학기술정책연구원
송위진 외 (2007) 脫추격형 기술혁신의 불확실성 대응 전략, 과학기술정책연구원
이승규 외 (2016.2) 2016년 KISTEP 10대 미래유망기술 선정에 관한 연구:
한국사회 삶의 만족과 사회적 신뢰를 향상시켜줄 미래 유망기술

해 외 문헌

- US OMB (2009.6) *DARPA Bridging the Gap Powered by Ideas*
David W. Cheney, Richard Van Att (2016.3) *DARPA 'S PROCESSES FOR
CREATING NEW PROGRAMS*, Technology Policy International, LLC
US NITRD (2011.12) *Future of Software Engineering Research*
Diomidis Spinellis (2016.11) *Research Priorities in the area of Software
Technologies*
Morency, John P. (2017) *Realities and Risks of Software-Defined Everything(SDx)*
Panetta, Kasey (2016.10.) *Gartner' s Top 10 Strategic Technology Trends for 2017*
Gartner (2017.10.4.) *Gartner Identifies the Top 10 Strategic Technology Trends
for 2018*
Khalil Rouhana (2016) *ICT in H2020 In support of the EU digital policies*

웹 사이트

- Guide2Research (www.guide2research.com)
IEEE Future Direction Initiatives (www.ieee.org/about/technologies/index.html)

Association for Computing Machinery(ACM) (www.acm.org)

Wikipedia (http://en.wikipedia.org/wiki/Analytic_hierarchy_process)

국가과학기술지식정보(NTIS) (www.ntis.go.kr)

US DARPA (www.darpa.mil)

US IARPA (www.iarpa.gov)

US NSF (www.nsf.gov)

EU H2020 (<http://cordis.europa.eu/>)

DBpia (<http://www.dbpia.co.kr>)

기타 참고자료와 그림 출처는 본문의 주석에 표시함

부 록

<부록 1> SW 분야 기술 수준: 2015년과 2016년 비교 (출처: IITP)

분류			기술격차(년)		기술수준(%)	
대 분 류	중분류	소분류	2016	2015	2016	2015
기 반 S W · 컴 퓨 팅	전체		1.8	2.0	71.7	74.1
	컴퓨팅 시스템	전체	2.1	2.0	75.5	76.6
		서버 (시스템)	2.3	2.4	72.3	72.6
		스토리지 (시스템)	1.7	1.6	78.7	80.6
		데이터센터	-	2.0	-	76.7
		클러스터·병렬 컴퓨팅	2.4	-	66.4	-
		기타 컴퓨팅 시스템	2.1	-	69.5	-
	High Performance 컴퓨팅	전체	-	2.8	-	66.3
		광·양자 컴퓨팅	-	3.0	-	63.2
		분산·병렬 컴퓨팅	-	2.6	-	67.4
		기타 HPC 기술	-	2.8	-	68.2
	인공지능·인지컴퓨팅	전체	2.2	2.4	73.9	70.5
		기계학습	-	2.5	-	70.4
		자연어처리	-	2.1	-	75.5
		인공지능	-	2.6	-	69.5
		휴먼컴퓨팅·마인드웨어	-	2.5	-	68.5
		기타 (2015)	-	2.6	-	68.6
		학습 및 추론	2.4	-	71.7	-
		상황이해	2.1	-	73.3	-
		언어이해	1.8	-	78.7	-
		시각이해	1.8	-	76.3	-
		인지컴퓨팅 (뉴로모픽 포함)	2.4	-	71.1	-
		기타 (2016)	2.2	-	71.7	-
	클라우드 컴퓨팅	전체	1.6	1.8	72.4	77.0
		클라우드 컴퓨팅 아키텍처	-	1.9	-	76.2
		클라우드 자원 가상화 및 통합	-	1.7	-	77.2
		클라우드 데이터, 저장, 분석	-	1.6	-	78.6
		기타 (2015)	-	1.8	-	75.9
		클라우드 플랫폼	1.8	-	73.7	-
		클라우드 연동	1.7	-	72.1	-
		클라우드 응용 및 서비스	1.5	-	76.0	-
		기타 (2016)	1.5	-	67.0	-
	모바일 컴퓨팅	전체	-	1.7	-	77.9
		센서 기반 컴퓨팅	-	1.9	-	75.9
		무선 컴퓨팅	-	1.5	-	80.7
		유비쿼터스/퍼베이시브 컴퓨팅	-	1.7	-	77.0
	빅데이터 플랫폼	전체	1.7	1.6	78.2	76.3
		데이터 수집	1.5	1.6	79.0	77.7
		저장·처리·관리	1.7	1.5	79.2	78.9
		분석 및 추론/예측	1.7	1.8	76.5	73.7
		활용 및 시각화 *서비스 및 응용	1.7	1.5	76.9	75.0
		기타 (2016)	1.7	-	79.6	-

S W	전체		1.9	1.9	76.7	76.8
	시스템 SW	전체	2.1	2.3	74.6	72.7
		운영체제	2.7	3.0	71.4	68.2
		스토리지 시스템	1.5	1.7	82.5	78.2
		DBMS	-	2.3	-	75.3
		가상화 SW	-	2.0	-	73.6
		프로그램 언어 및 개발도구	-	2.6	-	68.8
		SW공학도구	2.2	-	69.9	-
	미들웨어	전체	1.9	2.0	76.9	76.3
		분산/연계/통합 *분산시스템	2.0	2.0	76.0	76.0
		자원 관리	1.8	2.0	78.9	77.3
		서비스 플랫폼	1.9	2.0	76.1	75.6
	응용 SW	전체	1.8	1.3	78.7	81.5
		데스크톱 SW	-	1.6	-	78.6
		모바일 SW	-	0.9	-	87.5
		웹서비스	-	0.9	-	87.5
		클라우드 응용 SW	-	1.6	-	76.5
		빅데이터 응용 SW	-	1.8	-	74.5
		인터넷 응용 SW	-	1.0	-	85.0
		영상 처리 SW	1.9	-	80.5	-
		가상 시뮬레이션 SW	2.0	-	76.7	-
		자연어 처리	1.7	-	76.1	-
		음성 처리	1.6	-	82.1	-
		기업용 SW	1.8	-	78.1	-

〈부록 2〉 SW컴퓨팅산업원천기술개발사업 국가혁신형 과제(출처: NTIS)

- (엑소브레인-총괄) 사용자와 의사소통을 통한 지식공유 및 지능진화가 가능한 엑소브레인 SW 기술 개발 (ETRI 박상규/김현기, '13.5.1~'23.2.28, 1.0/1.0/1.0/1.0/1.0억원)
- (엑소브레인-1세부) 휴먼 지식증강 서비스를 위한 지능진화형 WiseQA 플랫폼 기술 개발 (ETRI 박상규/김현기, '13.5.1~'23.2.28, **1.0/35.0/35.0/32.7/32.7억원**)
- (엑소브레인-2세부) WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발 (솔트룩스 이경일, '13.5.1~'23.2.28, **1.0/23.0/23.0/21.5/21.5억원**)
- (엑소브레인-3세부) 비가호적 기법 기반 인간모사형 자가학습 지능 원천기술 개발 / 컨텍스트 인지형 Deep-Symbolic 하이브리드 지능 원천 기술 개발 및 언어 지식 자원 구축 (POSTECH 이근배 / KAIST 맹성현, '13.5.1~'23.2.28, 1.0/3.0/6.0/6.0/6.0억원)
- (엑소브레인-4세부) 자율지능형 지식/기기 협업 프레임워크 기술 개발 (ETRI 김성훈, '13.5.1~'23.2.28, **1.0/15.0/15.0억원**) *2015년 조기 종료
- (딥뷰-총괄) 대규모 실시간 영상 이해 기반의 시각지능 플랫폼 개발 (ETRI 박경/박종열, '14.4.1~'24.2.29, 1.0/1.0/1.0/1.0억원)
- (딥뷰-1세부) 실시간 대규모 영상 데이터 이해·예측을 위한 고성능 비주얼 디스커버리 플랫폼 개발 (ETRI 박경/박종열, '14.4.1~'24.2.29, **38.0/38.0/35.5/35.5억원**)
- (딥뷰-2세부) 대규모 실시간 비디오 분석에 의한 전역적 다중 관심객체 추적 및 상황 예측 기술 개발 (광주과학기술원 전문구, '14.4.1~'24.2.29, **20.0/20.0/18.7/18.7억원**)
- (딥뷰-3세부) 환경 재난·재해의 조기 감지를 위한 다형 영상 및 복합 센서 데이터 기반의 분석·예측 기술 개발 (계명대학교 박상원, '14.4.1~'24.2.29, **15.0/15.0억원**) *조기 종료
- (딥뷰-4세부) 예지형 시각 지능 원천 기술 개발 (서울대학교 이정무 / POSTECH 김대진, '14.4.1~'24.2.29, 3.0/6.0/6.0/6.0억원) *2차년도까지 경쟁으로 진행하다 통합
- (총괄) 빅데이터 처리 고도화 핵심 기술개발 사업 총괄 및 고성능 컴퓨팅 기술을 활용한 성능 가속화 기술 개발 (UNIST 김성엽, '15.6.1~'19.5.31, **30.00/28.09/24.30억원**)
- (1세부) 분산환경 인메모리 기술 기반의 복합형 고속 스트림 빅데이터 처리 기술 개발 (알티베이스 서광익, '15.6.1~'19.5.31, **24.00/22.40/19.38억원**)
- (2세부) 빅데이터 환경에서 비식별화 기법을 이용한 개인정보보호 기술 개발 (이지서티 이재훈, '15.6.1~'19.5.31, **16.00/15.00/12.99억원**)
- (BCI-총괄/1세부) 생각만으로 실생활 기기 및 AR/VR 디바이스를 제어하는 비침습 BCI 통합 뇌인지컴퓨팅 SW 플랫폼 기술 개발 (KIST 김래현, '17.4.1~'23.12.31, **30.00억원**)
- (BCI-2세부) 딥러닝을 이용하여 사람의 의도를 인지하는 BCI 기반 뇌인지컴퓨팅 기술 개발 (고려대학교 이성환, '17.4.1~'23.12.31, **20.00억원**)

〈부록 3〉 SW컴퓨팅산업원천기술개발사업 국가혁신형 과제(출처: NTIS)

○ 지능정보 플래그십 프로젝트 ('16~'20년, 750억원)

- (총괄/1세부) 자율지능 디지털 동반자 프레임워크 및 응용 연구 (KETI 정혜동, '17.4.1~'17.12.31, 50.0억원, ICT융합산업원천기술개발사업)
- 자율지능 동반자를 위한 적응형 기계학습 기술 연구 (KAIST 김종환, '16.12.1~'17.8.31(40.0억원)/'17.9.1~18.6.30(40.0억원), ICT융합산업원천기술개발사업) ※ 포항공대, 대구경북과학기술원, 위털루대, 알리(주) 참여
- 사용자의 의도와 맥락을 이해하는 지능형 인터랙션 기술 연구개발 (경북대 이민호, '16.12.1~'17.8.31(30.0억원)/'17.9.1~18.6.30(25.95억원), ※연세대, 한양대, KAIST, 시스트란, 이스트소프트 등 참여
- 대화 상대의 감성 추론 및 판단이 가능한 감성 지능 기술 연구 (KAIST 이수영, '16.12.1~'17.8.31(30.0억원)/'17.9.1~18.6.30(25.0억원), 첨단융복합콘텐츠기술개발사업) ※연세대, 한양대, 아크릴, 한국전자인증 참여

○ 인공지능 국가전략 프로젝트 ('17.7.1~'20.12.31, 사업단장 고려대학교 박재득, 일반사업)

- ※ 1차년도('17.7.1~18.3.31)
- 차세대 학습·추론 (37.5억원)
 - 의사결정 이유를 설명할 수 있는 인간 수준의 학습·추론 프레임워크 (UNIST 최재식, 18.75억원)
 - ※ 고려대, 연세의료원, AI트릭스 참여
 - 설명 가능한 인간 수준의 딥 기계학습 추론 프레임워크 개발 (KAIST 박현욱, 18.75억원)
- 비디오 튜링테스트 (42.0억원)
 - (총괄) 인간 수준의 비디오 이해 지능 및 검증 기술 개발
 - (1세부) 비디오 튜링 테스트를 통과할 수준의 비디오 스토리 이해 기반의 질의응답 기술 개발 (서울대 장병탁, 17.5억원) ※총괄과제 포함
 - (2세부) 비디오 이해를 위한 이벤트-상황 지식체계 학습 및 이벤트 인식/관계 추론 기술 개발 (KAIST 유창동, 연12.5억원)
 - (3세부) 비디오 이해를 위한 데이터 수집 및 보정 자동화 시스템 개발 (코난테크놀로지 윤덕호, 연12.0억원)

<부록 4> ICT 기술 분류 체계 (ITP, 2017.6.12.)

대분류	중분류	소분류
융합서비스 (Convergence Service)	사물인터넷 (Internet of Things)	IoT서비스 (Internet of Services)
		IoT플랫폼 (IoT Platform)
		IoT네트워킹 (IoT Networking)
		IoT디바이스 (IoT Device)
		사이버물리시스템(CPS) (Cyber Physical System)
	산업융합ICT (Industrial Convergence ICT)	농축수산ICT (Agriculture, Livestock & Fishing ICT)
		제조ICT (Manufacture ICT)
		에너지ICT (Energy ICT)
	공공융합ICT (Public Convergence ICT)	공공안전/재난예방ICT (Public Safety / Disaster Prevention ICT)
		의료ICT (Medical ICT)
		국방ICT (Defense ICT)
	스마트라이프ICT (Smart Life ICT)	스마트홈 (Smart Home)
		스마트시티 (Smart City)
		웰니스케어ICT (Wellness Care ICT)
	블록체인 (Block chain)	블록체인 핵심기술 (Blockchain Core technology)
		블록체인 플랫폼 (Blockchain Platform)
		블록체인 서비스 (Blockchain Service)
	기타융합서비스 (Other Convergence Service)	기타ICT융합서비스 (Other ICT Convergence Service)
이동통신 (Mobile Communication)	이동통신서비스 (Mobile Communication Service)	이동통신응용서비스 (Mobile Communication Applied Service)
		이동통신융합서비스 (Mobile Communication Converged Service)
	이동통신시스템 (Mobile Communication System)	셀룰러시스템무선접속기술 (Celluar System Wireless Access Technology)
		셀룰러시스템코어망기술 (Celluar System Core Network Technology)

	이동통신단말 및 부품 (Mobile Communication Device and Component)	근거리무선통신기술 (Local AreaWireless Communication Technology)
		응용및 융합시스템 기술 (Applied and Converged System Technology)
		이동통신단말기술 (Mobile Communication Device Technology)
		이동통신부품기술 (Mobile Communication Component Technology)
		이동통신시험/인증기술 (Mobile Communication Conformance/Certification Technology)
네트워크 (Network)	네트워크서비스 (Network Service)	네트워크 프로토콜 및 서비스 (Network Protocol and Service)
		네트워크 및 기능 가상화 응용서비스 (Application and Service based on Network and Function Virtualization)
		광통신 응용기술 (Optical Communications Application)
		기타 네트워크 서비스 (Other Network Services)
	네트워크플랫폼 (Network Platform)	네트워크운영체제 (Network Operating System)
		네트워크관리플랫폼 (Network Management Platform)
		미들박스 (Middle Box)
		네트워크 및 기능 가상화 서비스 플랫폼 (Network and Function Virtualization Services Platform)
		기타 네트워크 서비스 플랫폼 (OtherNetworkServicePlatforms)
	네트워크시스템 (Network System)	스위치/라우터 (Switch/Router)
		광전달망시스템 (Optical Transport Network System)
		광액세스망시스템 (Optical Access Network System)
		유무선융합네트워크시스템 (Fixed Mobile Convergence Network System)
		양자통신시스템 (Quantum Communications System)
	네트워크부품 (Network Device / Component)	스위치/라우터부품 (Switch/RouterDevice/Component)
		광통신부품 (Optical Communications Device/Component)
		양자통신부품 (Quantum Communications Device/Component)
방송·스마트미디어 (Broadcasting·Smartmedia)	방송·스마트미디어서비스 및 콘텐츠 (Broadcasting·Smartmedia Service and Content)	방송·스마트미디어서비스 (Broadcasting·Smartmedia Service)
		방송·스마트미디어콘텐츠 (Broadcasting·Smart media Content)
	방송·스마트미디어인프라 및 소비 (Broadcasting·Smartmedia for Infra and Consumption)	방송·스마트미디어매체 (Broadcasting·Smartmedia Infra)
		방송·스마트미디어단말 (Broadcasting·Smartmedia Terminal)

	방송·스마트미디어플랫폼 (Broadcasting·Smartmedia platform)	방송·스마트미디어이동방송 (Handheld Broadcasting·Smartmedia)
		방송·스마트미디어방통융합 (Convergence Broadcasting·Smartmedia)
		방송·스마트미디어실감방송 (Realistic Broadcasting·Smartmedia)
		방송·스마트미디어사이니지 (Broadcasting·Smartmedia for Digital Signage)
	방송·스마트미디어공공·복지 및 재난·안전 (Broadcasting·Smartmedia for Public Welfare and Disaster)	방송·스마트미디어공공복지 (Broadcasting·Smartmedia for Public Welfare)
		방송·스마트미디어재난·안전 (Broadcasting·Smartmedia for Disaster)
전파·위성 (Radiowave & Satellite)	전파자원·환경 (Radio Resource & Electromagnetic Environment)	스펙트럼공학 (Spectrum Engineering)
		전파전파/전자파해석 (Radiowave Propagation Modeling/EM Analysis)
		전자파적합성/전자파인체영향/ 전자파필스(EMC/EMF/EMP)
	전파기본·응용 (Radiowave Fundamental Technology&Applications)	안테나및전파기반디바이스 (Antenna&Radiowave Devices)
		에너지(전력)전송및수집 (Wireless Energy/Power Transfer&Harvesting)
		이미지·센서및계측 (Image/Sensor&RFMeasurement)
		정보전송 (Communication Technology for Information Transmission)
	위성·무인기 (Satellite&UAS)	위성HW/SW기술 (Satellite Hardware/Software Technology)
		위성서비스 (SatelliteService)
		무인기HW/SW기술 (UAS Hardware/SoftwareTechnology)
		무인기서비스 (UASService)
기반SW·컴퓨팅 (Infra SW and Computing)	인공지능 (Artificial Intelligence)	학습및추론기술 (Learning and Inference)
		상황이해기술 (Situational Understanding)
		언어이해기술 (Linguistic Understanding)
		시각이해기술 (Visual Understanding)
		인지컴퓨팅기술 (Cognitive Computing)
		기타인공지능기술 (Etc.)
	클라우드 (Cloud Computing)	클라우드플랫폼기술 (Cloud Platform)
		클라우드연동기술 (Cloud Interlocking)
		클라우드응용및서비스 (Cloud Applications and Services)

		기타클라우드기술 (Etc.)
	빅데이터 (BigData)	빅데이터수집기술 (Bigdata Collecting)
		빅데이터저장·처리·관리기술 (Bigdata Storing·Processing·Managing)
		빅데이터분석·예측기술 (Bigdata Analyzing·Prediction)
		서비스및응용 (Service and Application)
		기타빅데이터기술 (Etc.)
	컴퓨팅시스템 (Computing System)	서버시스템기술 (Server System)
		스토리지시스템기술 (Storage System)
		클러스터·병렬컴퓨팅기술 (Cluster·Parallel Computing)
		기타 컴퓨팅시스템 기술 (Etc.)
소프트웨어 (Software)	시스템 SW (System SW)	운영체제 (Operating System)
		데이터 관리SW (Data management SW)
		스토리지 SW (Storage SW)
		SW공학도구 (SW Engineering SW)
	미들웨어 (Middleware)	분산시스템SW (Decentralized system SW)
		자원 관리 (Resource Management)
		서비스 플랫폼 (Service Platform)
	응용SW (ApplicationSW)	영상 처리 SW (Image processing SW)
		가상 시뮬레이션 SW (Virtual Simulation SW)
		자연어 처리 (Natural language processing)
		음성 처리(Voice processing)
		기업용 SW(Enterprise SW)
정보보호 (Information security)	공통기반보안 (Fundamental security)	암호기술(Cryptography)
		인증/인가기술 (Authentication/Authorization)
		취약점분석및SW보안 (Vulnerability analysis and SW security)
	시스템·디바이스보안 (System·devicesecurity)	모바일·IoT디바이스보안 (Mobile·IoT devicesecurity)
		시스템보안 (System security)
		악성코드 분석 (Malware analysis)
	네트워크보안 (network security)	유선 네트워크 보안 (network(wired) security)
		무선 네트워크 보안 (wireless network security)
		보안분석 및 관제 (security management and monitoring)

	응용서비스 보안 (application security)	클라우드 보안 (cloud security)
		웹·이메일보안(Web·emailsecurity)
		데이터보안 (Data security)
		금융·핀테크 보안 (Financial security)
		디지털 포렌식 (Digital forensics)
	물리보안 (Physical security)	휴먼/바이오인식 (HumanIdentification/Biometrics)
		CCTV·무인전자감시 (CCTV·unmannedsurveillance)
	융합보안 (IT Convergence security)	스마트홈·시티보안 (Smarthome·citysecurity)
		산업제어시스템보안 (Industrial control system)
		지능형차량보안 (smart car security)
		헬스케어보안 (healthcare security)
		항공·조선해양·국방보안 (aircraft·ship building & offshore·military security)
		기타ICT융합보안(unclassified ICT convergence security)
디지털콘텐츠 (Digital contents)	실감형영상콘텐츠 (Realistic image/videocontents)	컴퓨터그래픽스 (Computergraphics)
		컴퓨터비전 (Computervision)
		3D/360°영상 (3Dstereoscopic/360°panoramicinage)
		플렌옵틱영상 (Plenopticimage)
		홀로그램 (Hologram)
	인터랙션 콘텐츠 (Interaction contents)	AR/MR (Augmentedreality/Mixedreality)
		VR (Virtualreality)
		NUI/NUX (Natural user interface / Naturaluserexperience)
		HCC (Human computer cooperation)
	감성콘텐츠 (Affective contents)	오감원천기술 (Fivesensesinformationprocessing)
		생체신호응용콘텐츠 (Bio-signal basedcontents)
		감성·인지 (Affective & cognitive contents)
	콘텐츠 유통/서비스 (Contents distribution/service)	콘텐츠분석·검색 (Contents analysis, contents search)
		콘텐츠보호·유통 (Contents protection, contents distribution)
		융복합응용서비스 (Contents fusion & convergence service)
		차세대플랫폼콘텐츠 (Next-generation platform based contents)
ICT디바이스	웨어러블디바이스	웨어러블HW플랫폼

(ICT Device)	(Wearable Device)	(Wearable Hardware Platform)
		웨어러블SW플랫폼 (Wearable Software Platform)
		기타웨어러블디바이스요소기술 (Other Essential Element Technologies for Wearable Device)
	ICT디바이스핵심부품및플랫폼 (Core Components & Platforms for ICT Devices)	지능형반도체 (Intelligent Semiconductor)
		스마트센서 및 센서플랫폼 (Smart Sensor and Sensor Platform)
		ICT디바이스 기타요소기술 (Other element technologies for the ICT Device)
	ICT융복합디바이스 (ICT Convergence Device)	3D프린터 (3D Printer)
		무인비행체 (Unmanned Aerial Vehicle, UAV)
		ICT지능형자동차 (ICT Smart Car)
		기타ICT융복합디바이스요소기술 (Other element technologies for the ICT Convergence Device)

<부록 5> ACM SIG 총 37개 (2017.12.20. 기준, ABC순)

SIGACCESS - Special Interest Group on Accessibility and Computing
SIGACT - Special Interest Group on Algorithms & Computation Theory
SIGAI - Special Interest Group on Artificial Intelligence
SIGAPP - Special Interest Group on Applied Computing
SIGARCH - Special Interest Group on Computer Architecture
SIGAda - Special Interest Group on Ada Programming Language
SIGBED - Special Interest Group on Embedded Systems
SIGBio - Special Interest Group on Bioinformatics, Computational Biology
SIGCAS - Special Interest Group on Computers and Society
SIGCHI - Special Interest Group on Computer-Human Interaction
SIGCOMM - Special Interest Group on Data Communication
SIGCSE - Special Interest Group on Computer Science Education
SIGDA - Special Interest Group on Design Automation
SIGDOC - Special Interest Group on Design of Communication
SIGEVO - Special Interest Group on Genetic and Evolutionary Computation
SIGGRAPH - Special Interest Group on Computer Graphics
SIGHPC - Special Interest Group on High Performance Computing
SIGIR - Special Interest Group on Information Retrieval
SIGITE - Special Interest Group on Information Technology Education
SIGKDD - Special Interest Group on Knowledge Discovery in Data
SIGLOG - Special Interest Group on Logic and Computation
SIGMETRICS - Special Interest Group on Measurement and Evaluation
SIGMICRO - Special Interest Group on Microarchitecture
SIGMIS - Special Interest Group on Management Information Systems
SIGMM - Special Interest Group on Multimedia Systems
SIGMOBILE - Special Interest Group on Mobility of Systems, Users, Data & Comp
SIGMOD - Special Interest Group on Management of Data
SIGOPS - Special Interest Group on Operating Systems
SIGPLAN - Special Interest Group on Programming Languages
SIGSAC - Special Interest Group on Security, Audit and Control
SIGSAM - Special Interest Group on Symbolic & Algebraic Manipulation
SIGSIM - Special Interest Group on Simulation
SIGSOFT - Special Interest Group on Software Engineering
SIGSPATIAL - Special Interest Group on Spatial Information
SIGUCCS - Special Interest Group on University & College Computing Services
SIGWEB - Special Interest Group on Hypertext, Hypermedia and Web
SIGecom - Special Interest Group on Electronic Commerce

〈부록 6〉 컴퓨터 과학 저널의 특집호(Special Issues) 주제

저널	주제
Software: Practice and Experience	Integration of Cloud, IoT and Big Data Analytics
Computers & Electrical Engineering	Edge-of-Things Computing for Smart Healthcare Systems: Opportunities and Challenges
Neurocomputing	Advances in Parallelism in Artificial Intelligence
Information Systems Journal	Indigenous Theory
Asian Journal of Control	Control Applications in Renewable Energy Systems
IET Radar, Sonar and Navigation	Cognitive Radar
IET Computer Vision	Computer Vision in Cancer Data Analysis
Optimization Letters	Applications of Distance Geometry
International Journal of Machine Learning and Cybernetics	Machine Learning Approaches and Challenges of Missing Data in the Era of Big Data
Journal of Computational Science	Computational Intelligence Paradigms in Recommender Systems and Online Social Networks
Computers and Security	Security and Privacy for Edge Computing: Current Progress and Future Challenge
Computers in Human Behavior	Digital interlocutors: theory and practice of interactions between human and machines
Future Generation Computer Systems	Accountability and Privacy Issues in Blockchain and Cryptocurrency
Future Generation Computer Systems	Cyber Threat Intelligence and Analytics
Future Generation Computer Systems	Autonomous Cloud
Future Generation Computer Systems	High Performance Services Computing and Internet Technologies
Future Generation Computer Systems	Edge of the Cloud
Computers and Industrial Engineering	Next Generation Smart Manufacturing and Service Systems using Big Data Analytics
Mobile Networks and Applications	Future Wireless Internet Technology and its Applications
<i>IEEE Computer Graphics and Applications</i>	<i>Visualization for Smart City Applications</i>
Decision Sciences	Information and Operational Decision Sciences: The Interplay of Information Technology and Operational Decision

	Sciences
<i>IEEE Journal on Selected Areas in Communications</i>	<i>Airborne Communication Networks</i>
Digital Signal Processing: A Review Journal	Co-operation and Joint Design of Communications and Radar Systems in a Crowded Spectrum
Computer Applications in Engineering Education	Innovations in Engineering Education with Digital Technologies
Pattern Recognition	Deep Learning for Computer Aided Cancer Detection and Diagnosis with Medical Imaging
Information Sciences	Distributed Event-Triggered Control and Estimation in Resource-Constrained Cooperative Networks
Pattern Recognition Letters	Graphonomics for e-citizens: e-health, e-society, e-education
Networks	Network based system optimization
Journal of Parallel and Distributed Computing	Cloud-of-Things and Edge Computing: Recent Advances and Future Trends
Sustainable Computing-Informatics & Systems	Recent Advances on Sustainability for green Cloud and Fog Computing
IET Communications	Recent Advances on 5G Communications
Pattern Recognition Letters	Learning Compact Representations for Scalable Visual Recognition and Retrieval
Journal of Parallel and Distributed Computing	Transmissible Cyber Threats in Distributed Systems
European Journal of Operational Research	Advances in Stochastic Optimization
AEÜ: International Journal of Electronics and Communications	Instrumentation Amplifiers: recent advances, design and implementation
Journal of Systems Architecture	"Embedded Systems in Applied Computing (SI-ESAC2018)
Knowledge-Based Systems	intelligent decision-making and consensus under uncertainty in inconsistent and dynamic environments
Displays	Advances in Smart Content-Oriented Display Technology
<i>IEEE Software</i>	<i>50 Years of Software Engineering</i>
<i>IEEE Wireless Communications</i>	<i>5G Testing and Field Trials</i>
<i>IEEE Transactions on Multimedia</i>	<i>Multimedia Big Data for Extreme Events</i>
Telematics and Informatics	ICTs and SUSTAINABLE Social Change (Code: ICTSSC)
Information Sciences	Advanced Methods for Evolutionary Many Objective Optimization
Future Generation Computer	"Novel edge computing-inspired approaches

Systems	and paradigms for mobile IoT applications”
Future Generation Computer Systems	Blockchain and Decentralization for Internet of Things
Signal Processing: Image Communication	Tensor Image Processing
<i>IEEE Network</i>	<i>Integration of Satellite and 5G Networks</i>
Robotics and Autonomous Systems	Autonomous Driving and Driver Assistance Systems
Swarm and Evolutionary Computation	Immune Computation and Applications
Information Fusion	Information Fusion for Emotion-aware Intelligent Systems
Displays	Visually Induced Motion Sensations
Computer Aided Geometric Design	Heat Diffusion Equation and Optimal Transport in Geometry Processing and Computer Graphics
Computers and Industrial Engineering	Data-driven decision making in supply chains
Journal of Knowledge Management	Knowledge Management in the Public Sector
Asian Journal of Control	SMC based observation, identification, uncertainties compensation and fault detection
Cognitive Systems Research	Innovative Dynamical Approaches to Cognitive Systems
Computers & Electrical Engineering	Towards Secure and Scalable Fog/Edge Computing
<i>IEEE Security and Privacy</i>	<i>Security and Privacy Research in Brazil</i>
International Journal of Human Computer Studies	Designing for human interactions with Cyber-Physical Systems
Differential Evolution	Swarm and Evolutionary Computation
Future Generation Computer Systems	Cloud and Fog Computing for Smart Cities
Pattern Recognition	Data Analytics and Visualisation
Computers and Graphics	Advances in Representation Learning
	Data-Driven Animation Techniques (D2AT)
Ad Hoc Networks	Recent advances on security and privacy in Intelligent Transportation Systems
Journal of Visual Communication and Image Representation	Emerging 3D and Immersive Data Processing and Evaluation Technologies
Computers & Electrical Engineering	Special Section on Soft Computing Approaches for Sustainable Systems
Internet Research	The Sharing Economy
European Journal of Information Systems	Information Systems and Innovation in the Public Sector
IET Software	Gamification and Persuasive Games for

	Software Engineering
Computers & Electrical Engineering	Artificial Intelligence and Robotics
Computers and Fluids	Fundamental and Engineering Aspects of Transfer Processes
Pattern Recognition Letters	Multiple-task Learning for Big Data (ML4BD)
Computers in Human Behavior	Anticipatory Computing: Crowd Intelligence from Social Network and Big Data
Pattern Recognition	Bio/Neuroscience inspired pattern recognition
Journal of Strategic Information Systems	Strategic Perspectives on Digital Work and Organizational Transformation
Neurocomputing	Deep Learning Neural Networks: Methods, Systems, and Applications
International Journal of Approximate Reasoning	Advances on Belief Functions and Their Applications
Journal of Systems and Software	Software Architecture
<i>IEEE Software</i>	<i>Cloudware Engineering</i>
<i>IEEE Software</i>	<i>The Social Developer: The Future of Software Development</i>
Pervasive and Mobile Computing	Mobile Crowdsourcing and Pervasive Computing for Smart Cities
Computers in Human Behavior	Emerging Technologies for Artifact Construction in Learning
Future Generation Computer Systems	Computation Intelligence for Energy Internet
Information Systems Journal	The Digitization of the Individual
Future Generation Computer Systems	Smart Data in Future Internet Technologies and Cloud Computing
Sustainable Computing-Informatics & Systems	Internet of Things for Efficient Energy Management of Computing Resources
Information Sciences	Privacy Computing: Principles and Applications
Electronic Commerce Research and Applications	ICT and the Sharing Economy: Economic Perspectives on the New E-Commerce
Signal Processing: Image Communication	Advances in statistical methods-based visual quality assessment
Journal of the Franklin Institute	Distributed Event-Triggered Control, Estimation, and Optimization
Pattern Recognition Letters	Learning and Recognition for Assistive Computer Vision
Sustainable Computing-Informatics & Systems	Sustainable Agriculture
Computers and Fluids	New computational methods for application of nanotechnology in engineering

Electronic Commerce Research and Applications	Recommender Systems with Side Information
Pattern Recognition Letters	Virtual Pattern Recognition Techniques for Non Verbal Human Behavior (NVHB)
Journal of Parallel and Distributed Computing	Exascale Applications and Software 2018
Journal of Computational Science	Soft Computing for Swarm Robotics: New Trends and Applications
Computers and Industrial Engineering	Digital and Organizational Transformation of Industrial Systems
Pattern Recognition Letters	"DLVA: Advances in Deep Learning and Visual Analytics for Intelligent Surveillance Systems"
Future Generation Computer Systems	Technological innovations in Digital transformation
Remote Sensing of Environment	Hyperspectral Remote Sensing of Vegetation Traits and Function
Electronic Commerce Research and Applications Tue 15 Jan	Consumer Brand Relationships in Virtual Markets
Information Systems Journal	Cybercrime and Computer 'Digital' Forensics

<부록 7> DARPA 현재 진행 219개 프로그램 (2017.8.20. 기준, ABC순)

100 Gb/s RF Backbone (100G)
Accelerated Computation for Efficient Scientific Simulation (ACCESS)
Active Authentication
Active Cyber Defense (ACD)
Active Interpretation of Disparate Alternatives (AIDA)
Adaptable Navigation Systems (ANS)
Adaptive Radar Countermeasures (ARC)
Adaptive RF Technology (ART)
Advanced RF Mapping (Radio Map)
Advanced Wide FOV Architectures for Image Reconstruction and Exploitation (AWARE)
Aerial Reconfigurable Embedded System (ARES)
Agile Teams (A-Teams)
Agnostic Compact Demilitarization of Chemical Agents (ACDC)
Airborne Launch Assist Space Access (ALASA)
Aircrew Labor In-Cockpit Automation System (ALIAS)
All Together Now (ATN)
Anomaly Detection at Multiple Scales (ADAMS)
Anti-Submarine Warfare (ASW) Continuous Trail Unmanned Vessel (ACTUV)
Arrays at Commercial Timescales (ACT)
Assured Arctic Awareness (AAA)
Assured Autonomy
Atomic Clock with Enhanced Stability (ACES)
Atoms to Product (A2P)
Autonomous Diagnostics to Enable Prevention and Therapeutics (ADEPT)
Battlefield Medicine
Behavioral Learning for Adaptive Electronic Warfare (BLADE)
Big Mechanism
Biochronicity
Biological Control
Biological Robustness in Complex Settings (BRICS)
Blue Wolf
Brandeis
Broad Operational Language Translation (BOLT)
Building Resource Adaptive Software Systems (BRASS)
Causal Exploration of Complex Operational Environments (Causal Exploration)
Circuit Realization at Faster Timescales (CRAFT)
Clean-slate design of Resilient, Adaptive, Secure Hosts (CRASH)
Collaborative Operations in Denied Environment (CODE)
Common Heterogeneous Integration and IP Reuse Strategies (CHIPS)

Communicating with Computers (CwC)
 Communications in Contested Environments (C2E)
 Communications Under Extreme RF Spectrum Conditions (CommEx)
 Complex Adaptive System Composition And Design Environment (CASCADE)
 Computational Leverage Against Surveillance Systems (CLASS)
 Computational Simulation of Online Social Behavior (SocialSim)
 CONverged Collaborative Elements for RF Task Operations (CONCERTO)
 Cross-Domain Maritime Surveillance and Targeting (CDMaST)
 Cyber Assured Systems Engineering (CASE)
 Cyber Fault-tolerant Attack Recovery (CFAR)
 Cyber Grand Challenge (CGC)
 DAHI - Electronic-Photonic Heterogeneous Integration (E-PHI)
 DAHI - Foundry Technology
 Data-Driven Discovery of Models (D3M)
 Deep Exploration and Filtering of Text (DEFT)
 Deep Purposeful Learning (Deep Purple)
 Dialysis-Like Therapeutics (DLT)
 Direct On-Chip Digital Optical Synthesizer (DODOS)
 Direct Sampling Digital Receiver (DISARMER)
 Dispersed Computing
 Distributed Agile Submarine Hunting (DASH)
 Distributed Battle Management (DBM)
 Diverse Accessible Heterogeneous Integration (DAHI)
 Dynamic Network Adaptation for Mission Optimization (DyNAMO)
 Edge-Directed Cyber Technologies for Reliable Mission Communication (EdgeCT)
 Efficient Linearized All-Silicon Transmitter ICs (ELASTx)
 Electrical Prescriptions (ElectRx)
 Enabling Quantification of Uncertainty in Physical Systems (EQUiPS)
 ENGAGE
 Engineered Living Materials (ELM)
 Enhanced Attribution
 EXCALIBUR
 Experimental Spaceplane
 Explainable Artificial Intelligence (XAI)
 Extended Solids (XSolids)
 Extreme DDoS Defense (XD3)
 Extreme Optics and Imaging (EXTREME)
 Fast Lightweight Autonomy (FLA)
 Folded Non-Natural Polymers with Biological Function (Fold Fx)
 Fundamental Limits of Learning (FunLoL)
 Fundamental Limits of Photon Detection (Detect)

Fundamentals of Complex Collectives (FunCC)
Gamifying the Search for Strategic Surprise (GS3)
Gremlins
Ground Truth
Ground X-Vehicle Technologies (GXV-T)
Hallmark
Hand Proprioception and Touch Interfaces (HAPTIX)
Harnessing Autonomy for Countering Cyberadversary Systems (HACCS)
Hierarchical Identify Verify Exploit (HIVE)
High power Amplifier using Vacuum electronics for Overmatch Capability (HAVOC)
High-Assurance Cyber Military Systems (HACMS)
Hunter
Hydra
Hyper-wideband Enabled RF Messaging (HERMES)
Hypersonic Air-breathing Weapon Concept (HAWC)
Improv
In Vivo Nanoplatfoms (IVN)
Inbound, Controlled, Air-Releasable, Unrecoverable Systems (ICARUS)
Insect Allies
Insight
Integrity and Reliability of Integrated Circuits (IRIS)
Intense and Compact Neutron Sources (ICONS)
INTERfering and Co-Evolving Prevention and Therapy (INTERCEPT)
Intrachip/Interchip Enhanced Cooling (ICECool)
Lagrange
Laser UV Sources for Tactical Efficient Raman (LUSTER)
Leveraging the Analog Domain for Security (LADS)
Living Foundries
Local Control of Materials Synthesis (LoCo)
Low Cost Thermal Imager - Manufacturing (LCTI-M)
Low Resource Languages for Emergent Incidents (LORELEI)
Magnetic Miniaturized and Monolithically Integrated Components (M3IC)
MAGneto Hydrodynamic Explosive Munition (MAHEM)
Make-It
Materials Development for Platforms (MDP)
Materials for Transduction (MATRIX)
Media Forensics (MediFor)
Memex
Mesodynamic Architectures (Meso)
Microphysiological Systems (MPS)
Microscale Plasma Devices (MPD)

Military Imaging and Surveillance Technology (MIST)
Mining and Understanding Software Enclaves (MUSE)
Mission-oriented Resilient Clouds (MRC)
Mobile Hotspots
Mobile Offboard Clandestine Communications and Approach (MOCCA)
Modeling Adversarial Activity (MAA)
Models, Dynamics and Learning (MoDyL)
Modular Optical Aperture Building Blocks (MOABB)
Molecular Informatics
Multi-Azimuth Defense Fast Intercept Round Engagement System (MAD-FIRES)
Multifunction RF (MFRF)
Near Zero Power RF and Sensor Operations (N-ZERO)
Neural Engineering System Design (NESD)
Neuro Function, Activity, Structure, and Technology (Neuro-FAST)
Next Generation Social Science (NGS2)
Nitride Electronic NeXt-Generation Technology (NEXT)
Open Manufacturing
OrbitOutlook
Pandemic Prevention Platform (P3)
Pathogen Predators
Phoenix
Photonically Optimized Embedded Microprocessors (POEM)
Pixel Network for Dynamic Visualization (PIXNET)
Plan X
Positioning System for Deep Ocean Navigation (POSYDON)
Power Efficiency Revolution for Embedded Computing Technologies (PERFECT)
Precise Robust Inertial Guidance for Munitions (PRIGM)
Probabilistic Programming for Advancing Machine Learning (PPAML)
Program in Ultrafast Laser Science and Engineering (PULSE)
Prometheus
Prophecy (Pathogen Defeat)
Prototype Resilient Operations Testbed for Expeditionary Urban Operations (PROTEUS)
Quantitative Crisis Response (QCR)
Quantum Orbital Resonance Spectroscopy (QORS)
Quantum-Assisted Sensing and Readout (QuASAR)
Quiness
RadioBio
Rapid Attack Detection, Isolation and Characterization Systems (RADICS)
Rapid Threat Assessment
Reactive Material Structures (RMS)
Reliable Neural-Interface Technology (RE-NET)

Resilient Synchronized Planning and Assessment for the Contested Environment (RSPACE)
Restoring Active Memory (RAM)
Retrodirective Arrays for Coherent Transmission (ReACT)
Revolutionary Enhancement of Visibility by Exploiting Active Light-fields (REVEAL)
Revolutionizing Prosthetics
RNET
Robotic Servicing of Geosynchronous Satellites (RSGS)
Robust Automatic Transcription of Speech (RATS)
Safe Genes
SAFEWARE
Seeker Cost Transformation (SECTR)
Semiconductor Technology Advanced Research Network (STARnet)
Shared Spectrum Access for Radar and Communications (SSPARC)
SIGMA
Signal Processing at RF (SPAR)
Simplifying Complexity in Scientific Discovery (SIMPLEX)
Small Satellite Sensors
Soldier Centric Imaging via Computational Cameras (SCENICC)
Soldier Protection Systems (SPS)
Space Enabled Effects for Military Engagements (SeeMe)
Space Surveillance Telescope (SST)
Space/Time Analysis for Cybersecurity (STAC)
Spatial, Temporal, and Orientation Information in Contested Environments (STOIC)
Spectral Combs from UV to THz (SCOUT)
Spectrum Collaboration Challenge (SC2)
Squad X Core Technologies (SXCT)
Strategic Social Interaction Modules (SSIM)
Supply Chain Hardware Integrity for Electronics Defense (SHIELD)
Synergistic Discovery and Design (SD2)
System of Systems Integration Technology and Experimentation (SoSITE)
Systems-Based Neurotechnology for Emerging Therapies (SUBNETS)
Tactical Boost Glide (TBG)
Tactical Undersea Network Architectures (TUNA)
Tailorable Feedstock and Forming (TFF)
Target Recognition and Adaption in Contested Environments (TRACE)
Targeted Neuroplasticity Training (TNT)
Technologies for Host Resilience (THoR)
Tern
THz Electronics
Topological Excitations in Electronics (TEE)

Transformative Design (TRADES)
Transmit and Receive Optimized Photonics (TROPHY)
Transparent Computing
Trusted Integrated Circuits (TRUST)
Unconventional Processing of Signals for Intelligent Data Exploitation (UPSIDE)
Vanishing Programmable Resources (VAPR)
Vertical Takeoff and Landing Experimental Plane (VTOL X-Plane)
Vetting Commodity IT Software and Firmware (VET)
Video Synthetic Aperture Radar (ViSAR)
Wafer Scale Infrared Detectors (WIRED)
Warfighter Analytics using Smartphones for Health (WASH)
Warrior Web
Wireless Network Defense
World Modelers
Wound Stasis System (WSS)
XDATA
Z-Man

<부록 8> IARPA 현재 진행 35개 중 주요 프로그램 (2017.8.20. 기준, ABC 순)

Program	Research Area	Program Manager
Aladdin Video	Image, photograph, video, multimedia, computer vision, natural language processing, image processing, big data, video analytics, machine learning, speech processing	Terry Adams
Amon-Hen	Space situational awareness, optical interferometry, fiber optics, image reconstruction, computational imaging, small aperture telescopes, adaptive optics, astrometry, astronomy, optical sensors, optics, optical design	Merrick J. DeWitt
Babel	Multilingual/multidialectal speech recognition, keyword search algorithms, speech recognition in noisy environments, low resource languages, rapid adaptation to new languages and environments, machine learning	Carl Rubino
C3	Advanced/alternative computing technologies, superconducting microelectronics	Marc Manheimer
CAUSE	Cybersecurity, cyber-event forecasting, cyber-actor behavior and cultural understanding, threat intelligence, threat modeling, cyber-event coding, cyber-kinetic event detection	Robert Rahmer
CORE3D	Multi-view satellite image processing, multi-modal information fusion, deep learning, remote sensing, photogrammetry, image segmentation and classification, multispectral imagery processing, and geospatial volumetric 3D data representation methods	Hakjae Kim
CREATE	Logic, human judgment, critical thinking, forecasting	Steven Rieber
DIVA	Machine learning, deep learning or hierarchical modeling, artificial intelligence, object detection, recognition, person detection and re-identification, person action recognition, video activity detection, tracking across multiple non-overlapping camera viewpoints, 3D reconstruction from video, super-resolution, stabilization, statistics, probability and mathematics	Terry Adams
EELIX	Biological detection, systems biology, synthetic biology, genome editing, bioinformatics, evolutionary biology	Amanda Dion-Schultz
Finder	Geolocation, localization, geospatial fusion, data fusion, machine learning, big data, image processing, image, photograph, video, multimedia, computer vision, natural language processing	Hakjae Kim
Fun_GCAT	Bioinformatics, DNA sequence screening, functional genomics, systems biology, infectious disease, and synthetic biology	John Julius
EUSE	Technical emergence, text analytics, knowledge discovery, big data, social network analysis, natural language processing, forecasting, machine learning	Robert Rahmer
HECTOR	Secure multiparty computation, homomorphic encryption, verifiable computing, compilers, programming languages, automated security analysis	Mark Heiligman
HEC	Forecasting, human judgment, machine learning, decision making, human/machine interfaces, text analysis	Seth Goldstein
HFGeo	Communication systems, ionosphere, antennas, geolocation, electromagnetics, radio frequency	Torreon Creekmore
Ithildin	Sorbent chemistry, polymer chemistry, encapsulation, nanotechnology, micro-engineered materials, reaction kinetics, chemical analysis techniques	Kristy DeWitt
Janus	Computer vision, image processing, pattern recognition, biometrics, facial recognition, identity intelligence, computer graphics	Chris Boehnen
LoqiQ	Advanced/alternative computing platforms, quantum information sciences, qubits systems	Brad Blakestad
MAEGLIN	Liquid and gas chromatography, flow cytometry, ion mobility spectrometry, preconcentrators and sorbents, ionization techniques, mass spectrometry, MEMS technology, microfluidics, computational fluid dynamics, signature library construction and use, clutter detection, multicomponent fits, basis set transformations, micro vacuum pump technology, low power electronics, device SWAP optimization, regenerative energy sources and energy harvesting techniques.	Kristy DeWitt
MATERIAL	Natural language processing, machine translation, cross-lingual information retrieval, domain recognition and adaptation, multilingual ontologies, Multilingual speech recognition, cross-lingual summarization, keyword search algorithms, low resource languages, automatic language identification, machine learning, rapid adaptation to new languages, domains and genres	Carl Rubino
Mercury	SIGINT analytics, event forecasting, machine learning, streaming data, data fusion, disease outbreak	Kristen Jordan
MICrONS	Theoretical neuroscience , computational neuroscience , machine learning, connectomics, brain activity mapping	David Markowitz
MOSAIC	Behavioral science, cognitive psychology, human performance, mobile computing, context sensing, signal processing, data fusion, machine learning, data privacy and security	Alexis Jeannotte
Ordin	Biometrics, presentation attack, machine learning, computer vision	Chris Boehnen
Proteos	Proteomics, single amino acid polymorphisms, genetically variable peptides, genomics, statistical analysis, molecular biology, human identification, forensic analysis, trace evidence samples	Kristen Jordan
QEO	Quantum annealing, combinatorial optimization, quantum error correction, non-stoquastic quantum interactions, multi-spin entanglement, classical annealing, Quantum Monte Carlo, population annealing, parallel tempering, adaptive annealing schedules with measurement feedback	Karl Roenigk
RAVEN	Microelectronics, nondestructive analysis, nanoscale imaging, hardware assurance	Carl McCants
SCITE	Engineering enterprises that detect low probability events with low accuracy sensors, innovative research methods to evaluate analytic and forecasting tradecraft, innovative statistical methods to estimate performance of systems addressing complex analysis and forecasting problems, scientific research on organizational lessons-learned methods, evidence-based forecasting methods, inductive logic, probabilistic reasoning and its application to analytic tradecraft	Paul Lehner
SHARP	Cognition , psychometrics , fluid reasoning and intelligence , neuroscience , human performance	Alexis Jeannotte
SILMARILS	Chemical detection and identification (including standoff, remote, and ultra-compact/low power approaches), spectroscopy/spectrometry/chromatography, optical sensors, novel laser designs, frequency combs, nonlinear optics, fiber optic sensors/lasers/devices	Kristy DeWitt
SLICE	Communication systems, geolocation, electromagnetics, radio frequency	Jeff Weinschenk
SuperTools	Superconducting electronics, advancements/alternatives to semiconductor-based exo-scale computing, cryogenic computing	Mark Heiligman
TIC	Cybersecurity and information assurance, hardware assurance, microelectronics	Carl McCants
UnderWatch	Autonomous Underwater Vehicles (AUVs), associated technology enablers	Jeff Weinschenk
VirtUE	Computer virtualization, operating systems, cyber security, vulnerability analysis, insider threat remediation and detection, active defense, big data analytics, sensor fusion, user interfaces, anomalous event detection	Kerry Long

<부록 9> CISE의 조직별 사업 [범례: C(Crosscutting), N(NSF-wide)]

- Division of Computing and Communication Foundations (CCF)
 - Computing and Communication Foundations (CCF): Core Programs
 - Algorithmic Foundations (AF)
 - Communications and Information Foundations (CIF)
 - Software and Hardware Foundations (SHF)
 - Additional Funding Opportunities for the CCF Community
 - Algorithms in the Field (AitF)
 - CISE Research Infrastructure (CRI)
 - CISE-MPS Interdisciplinary Faculty Program in Quantum Information Science
 - Computer and Information Science and Engineering (CISE) Research Initiation Initiative (CRII)
 - Computer Science for All (CSforAll:RPP) C
 - Critical Techniques, Technologies and Methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering (BIGDATA)
 - Cyber-Physical Systems (CPS)
 - Designing Materials to Revolutionize and Engineer our Future (DMREF)
 - Energy-Efficient Computing: from Devices to Architectures (E2CDA)
 - Expeditions in Computing
 - Exploiting Parallelism and Scalability (XPS)
 - FAILURE-RESISTANT SYSTEMS (FRS)
 - NSF/Intel Partnership on Computer Assisted Programming for Heterogeneous Architectures (CAPA)
 - NSF/Intel Partnership on Visual and Experiential Computing (VEC)
 - Scalable Parallelism in the Extreme (SPX)
 - Secure and Trustworthy Cyberspace (SaTC)
 - Smart and Autonomous Systems (S&AS)
 - Smart and Connected Communities (S&CC)
 - Smart and Connected Health (SCH)
 - Software Infrastructure for Sustained Innovation (SSE, SSI, S2I2) NC
 - Training-based Workforce Development for Advanced Cyberinfrastructure (CyberTraining)
 - Transdisciplinary Research in Principles of Data Science Phase I (TRIPODS) C
 - United States-Israel Collaboration in Computer Science (USICCS)
 - US-Japan Big Data and Disaster Research (BDD)
 - Wireless Innovation between Finland and US (WiFiUS)
 - Connections in Quantum Information Science (CQIS) C
 - Ideas Lab: Practical Fully-Connected Quantum Computer Challenge (PFCQC)
 - IUSE / Professional Formation of Engineers: REvolutionizing engineering and computer science Departments (IUSE/PFE: RED)
 - Semiconductor Synthetic Biology for Information Processing and Storage Technologies (SemiSynBio)

- Division of Computer and Network Systems (CNS)
 - Computer and Network Systems (CNS): Core Programs
 - Computer Systems Research (CSR)
 - Networking Technology and Systems (NeTS)
 - Education and Workforce Program
 - Broadening Participation in Computing Alliance Program (BPC-A)
 - Research Infrastructure Program
 - CISE Research Infrastructure (CRI)
 - Global Environment for Networking Innovations (GENI)
 - Major Research Instrumentation Program: (MRI) N
 - Additional Funding Opportunities for the CNS Community
 - ADVANCE: Increasing the Participation and Advancement of Women in Academic Science and Engineering Careers N
 - Algorithms in the Field (AitF)
 - Campus Cyberinfrastructure (CC*)
 - CISE Research Infrastructure: Mid-Scale Infrastructure - NSFCloud (CRI: NSFCloud)
 - Computer and Information Science and Engineering (CISE) Research Initiation Initiative (CRII)
 - Computer Science for All (CSforAll:RPP) C
 - Critical Techniques, Technologies and Methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering (BIGDATA)
 - Cyber-Physical Systems (CPS)
 - Designing Materials to Revolutionize and Engineer our Future (DMREF)
 - Expeditions in Computing
 - Exploiting Parallelism and Scalability (XPS)
 - Future Internet Architectures -- Next Phase (FIA-NP)
 - Graduate Research Fellowship Program (GRFP) N
 - Industry-University Cooperative Research Centers Program (IUCRC) N
 - Japan-US Network Opportunity 2 (JUNO)
 - NSF/Intel Partnership on Cyber-Physical Systems Security and Privacy (CPS-Security)
 - NSF/Intel Partnership on Visual and Experiential Computing (VEC)
 - Partnerships for Innovation: Building Innovation Capacity (PFI:BIC)
 - Research Experiences for Undergraduates (REU) N
 - Scalable Parallelism in the Extreme (SPX)
 - Secure and Trustworthy Cyberspace (SaTC)
 - Secure and Trustworthy Cyberspace: Secure, Trustworthy, Assured and Resilient Semiconductors and Systems (SaTC: STARSS)
 - Smart and Autonomous Systems (S&AS)
 - Smart and Connected Communities (S&CC)
 - Smart and Connected Health (SCH)
 - Software Infrastructure for Sustained Innovation (SSE, SSI, S2I2) NC
 - US Ignite:
 - US-Japan Big Data and Disaster Research (BDD)
 - Wireless Innovation between Finland and US (WiFiUS)

- Critical Resilient Interdependent Infrastructure Systems and Processes FY17 (CRISP)
- Innovation Corps (I-Corps TM) - National Innovation Network Nodes Program (I-Corps™ Nodes)
- IUSE / Professional Formation of Engineers: REvolutionizing engineering and computer science Departments (IUSE/PFE: RED)
- NSF/Intel Partnership on Information-Centric Networking in Wireless Edge Networks (ICN-WEN)
- NSF/VMware Partnership on Software Defined Infrastructure as a Foundation for Clean-Slate Computing Security (SDI-CSCS)
- Optics and Photonics (OP) C
- Platforms for Advanced Wireless Research (PAWR): Establishing the PAWR Project Office (PPO) (PAWR/PPO)
- Spectrum Efficiency, Energy Efficiency, and Security (SpecEES): Enabling Spectrum for All C
- Tomorrow's Internet Project Office (TIPOFF)

□ Division of Information & Intelligent Systems (IIS)

- Information and Intelligent Systems (IIS): Core Programs
 - Cyber-Human Systems (CHS)
 - Information Integration and Informatics (III)
 - Robust Intelligence (RI)
- Additional Funding Opportunities for the IIS Community
 - Algorithms in the Field (AitF)
 - CISE Research Infrastructure (CRI)
 - Collaborative Research in Computational Neuroscience (CRCNS)
 - Computer and Information Science and Engineering (CISE) Research Initiation Initiative (CRII)
 - Computer Science for All (CSforAll:RPP) C
 - Critical Techniques, Technologies and Methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering (BIGDATA)
 - Cyber-Physical Systems (CPS)
 - Cyberlearning and Future Learning Technologies C
 - Designing Materials to Revolutionize and Engineer our Future (DMREF)
 - Documenting Endangered Languages (DEL)
 - Expeditions in Computing
 - Exploiting Parallelism and Scalability (XPS)
 - National Robotics Initiative 2.0: Ubiquitous Collaborative Robots (NRI-2.0)
 - NSF/Intel Partnership on Visual and Experiential Computing (VEC)
 - Partnerships for Innovation: Building Innovation Capacity (PFI:BIC)
 - Scalable Parallelism in the Extreme (SPX)
 - Secure and Trustworthy Cyberspace (SaTC)
 - Smart and Autonomous Systems (S&AS)
 - Smart and Connected Communities (S&CC)
 - Smart and Connected Health (SCH)

- Software Infrastructure for Sustained Innovation (SSE, SSI, S2I2) NC
- US Ignite:
- US-Japan Big Data and Disaster Research (BDD)
- IUSE / Professional Formation of Engineers: REvolutionizing engineering and computer science Departments (IUSE/PFE: RED)

□ Office of Advanced Cyberinfrastructure (OAC)

- Academic Research Infrastructure Program: Recovery and Reinvestment (ARI-R²) C
- Benchmarks of Realistic Scientific Application Performance of Large-Scale Computing Systems (BRAP)
- CISE Research Infrastructure: **Mid-Scale Infrastructure** - NSFCLOUD (CRI: NSFCLOUD)
- Computational and Data-Enabled Science and Engineering (CDS&E)
- Computational and Data-Enabled Science and Engineering in Mathematical and Statistical Sciences (CDS&E-MSS)
- Computer and Information Science and Engineering (CISE) Research Initiation Initiative (CRII)
- Critical Resilient Interdependent Infrastructure Systems and Processes FY17 (CRISP)
- Cyber-Physical Systems (CPS)
- Cyberinfrastructure for Emerging Science and Engineering Research (CESER)
- Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21) NC
- Cybersecurity Innovation for Cyberinfrastructure (CICI)
- Designing Materials to Revolutionize and Engineer our Future (DMREF)
- Division of Advanced Cyberinfrastructure Programs (ACI)
 - Campus Cyberinfrastructure (CC*)
 - Data Infrastructure Building Blocks (DIBBs)
 - International Research Network Connections (IRNC)
 - Petascale Computing Resource Allocations (PRAC)
 - Software Development for Cyberinfrastructure (SDCI)
 - Software Infrastructure for Sustained Innovation (SSE, SSI, S2I2) NC
 - Towards a Leadership-Class Computing Facility - Phase 1
 - Virtual Organizations as Sociotechnical Systems (VOSS)
- EarthCube:
- Expeditions in Computing
- Exploiting Parallelism and Scalability (XPS)
- Explosives and Related Threats: Frontiers in Prediction and Detection (EXP) C
- Implementation of NSF CIF21 Software Vision (SW-Vision) C
- IUSE / Professional Formation of Engineers: REvolutionizing engineering and computer science Departments (IUSE/PFE: RED)
- Network for Computational Nanotechnology (NCN)
- Research Experiences for Undergraduates (REU) N
- Scalable Parallelism in the Extreme (SPX)
- Science, Engineering and Education for Sustainability NSF-Wide Investment (SEES) NC
 - Dimensions of Biodiversity FY2017
- Secure and Trustworthy Cyberspace (SaTC)

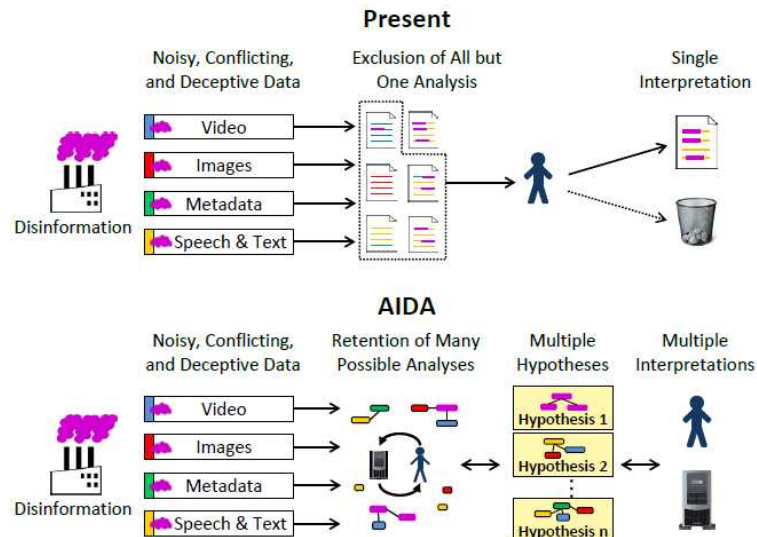
- Tomorrow's Internet Project Office (TIPOFF)
- Training-based Workforce Development for Advanced Cyberinfrastructure (CyberTraining)

<부록 10> DARPA 예비 후보

[DARPA 예비후보1] Active Interpretation of Disparate Alternatives (AIDA)

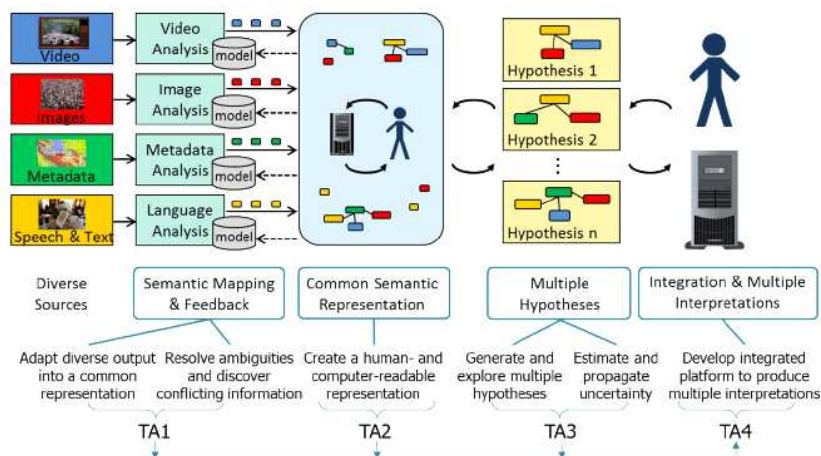
PM: DR. BOYAN ONYSHKEVYCH ※ '17.4.3 공지, '17.6.6 제안서 접수 54개월 수행

다양한 비정형화된 데이터를 동시에 표현하고 분석하여 사건, 상황, 트렌드 등에 대한 복수 가설을 설정하여 의사결정자를 지원하는 엔진 개발



The goal of Active Interpretation of Disparate Alternatives (AIDA) is to develop a multihypothesis semantic engine that generates explicit alternative interpretations of events, situations, and trends from a variety of unstructured sources, for use in noisy, conflicting, and potentially deceptive information environments.

- TA1 - Semantic Mapping and Feedback
- TA2 - Common Semantic Representation
- TA3 - Multiple Hypotheses
- TA4 - Integration and Multiple Interpretations
- TA5 - Data



TA1-TA3 복수 경쟁 수행 예상, TA4-TA5 단일 수행 예상

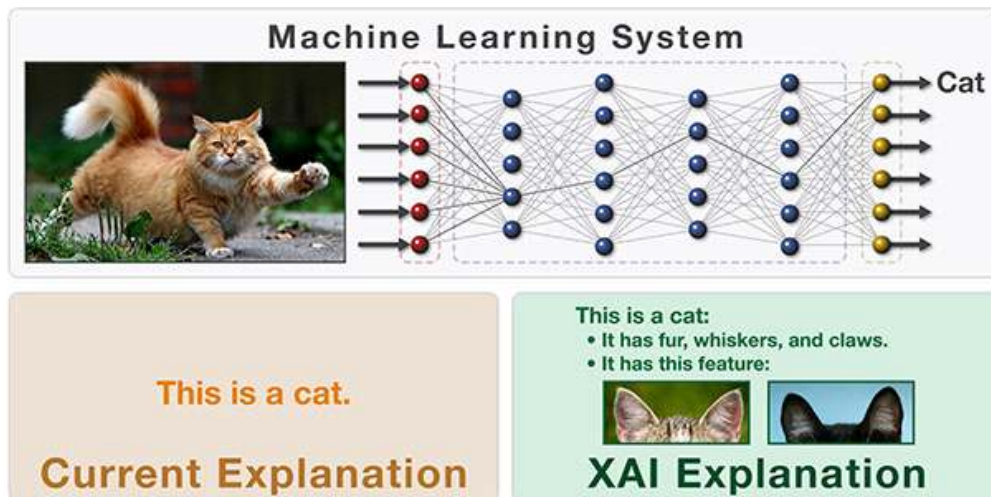
*국내 과제 기획(예: AI 국가전략 프로젝트)에서 자주 언급되었으나 실제 과제화되지 못함

*TA1의 여러 분석의 준비도, 특히 시각 분석의 기술 수준 확인 필요

[DARPA 예비후보2] Explainable Artificial Intelligence (XAI)

PM: MR. DAVID GUNNING ※'17.5~'21.4 (48개월)

현재 결과만 도출해서 추천하는 딥러닝의 한계를 극복하고 추천한 이유를 설명할 수 있도록 기계학습 테크닉 개발 (단, 멀티미디어 데이터로부터의 관심 이벤트 분류, 자율기계의 의사결정 문제에만 국한하여 적용)



The Explainable AI (XAI) program aims to create a suite of machine learning techniques that:

- Produce more explainable models, while maintaining a high level of learning performance (prediction accuracy); and
- Enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.

New machine-learning systems will have the ability to explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future. The strategy for achieving that goal is to develop new or modified machine-learning techniques that will produce more explainable models. These models will be combined with state-of-the-art human-computer interface techniques capable of translating models into understandable and useful explanation dialogues for the end user. Our strategy is to pursue a variety of techniques in order to generate a portfolio of methods that will provide future developers with a range of design options covering the performance-versus-explainability trade space.

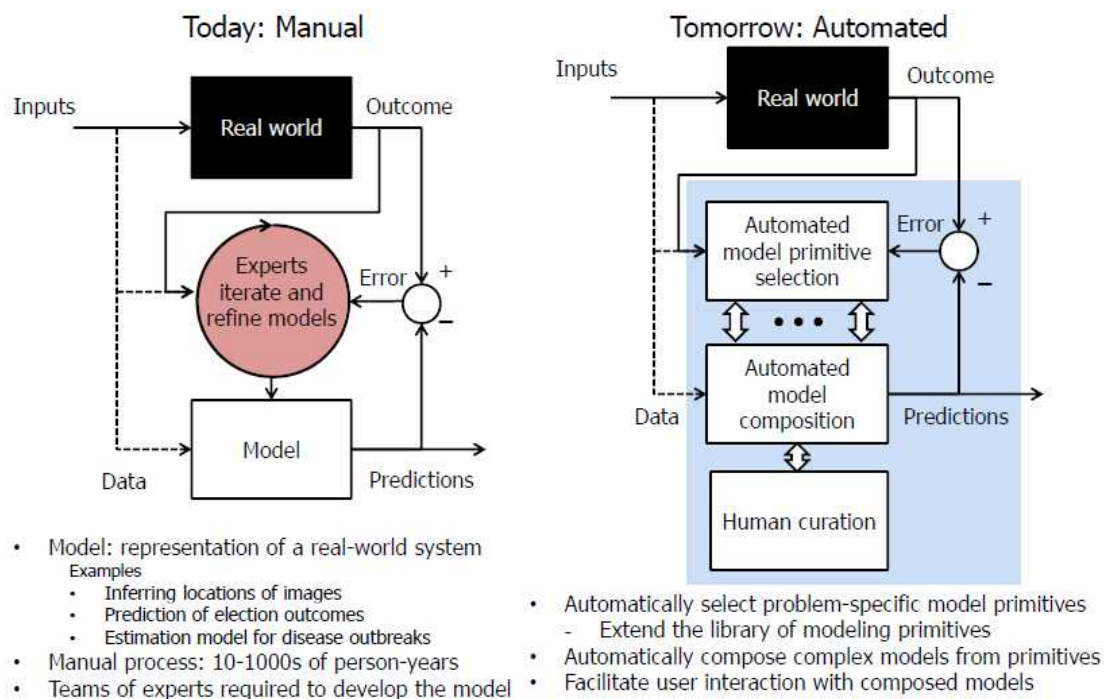
The XAI program will focus the development of multiple systems on addressing challenges problems in two areas: (1) **machine learning problems to classify events of interest in heterogeneous, multimedia data**; and (2) **machine learning problems to construct decision policies for an autonomous system to perform a variety of simulated missions**. These two challenge problem areas were chosen to represent the intersection of two important machine learning approaches (classification and reinforcement learning) and two important operational problem areas for the Department of Defense (intelligence analysis and autonomous systems).

[DARPA 예비후보3] Data-Driven Discovery of Models (D3M)

PM: MR. WADE SHEN ※'16.6.15 공지, 48개월 2022년 종료

데이터 과학 지식이 부족한 도메인 전문가들이 실증 연구를 할 수 있는 자동화된 모델 발견 시스템 개발

The Data-Driven Discovery of Models (D3M) program aims to **develop automated model discovery systems that enable users with subject matter expertise but no data science background to create empirical models of real, complex processes**. This capability will enable subject matter experts to create empirical models without the need for data scientists, and will increase the productivity of expert data scientists via automation. The D3M automated model discovery process, depicted in the figure, will be enabled by three key technologies to be developed in the course of the program:

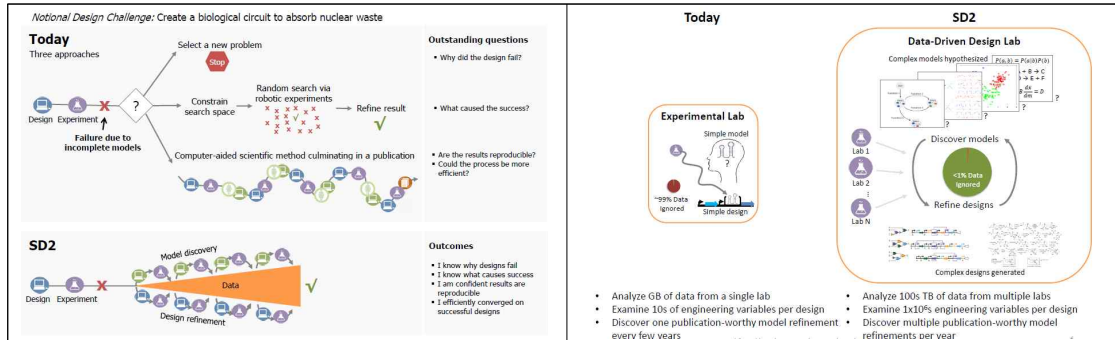


- **TA1: A library of selectable primitives.** A discoverable archive of data modeling primitives will be developed to serve as the basic building blocks for complex modeling pipelines.
- **TA2: Automated composition of complex models.** Techniques will be developed for automatically selecting model primitives and for composing selected primitives into complex modeling pipelines based on user-specified data and outcome(s) of interest.
- **TA3: Human-model interaction** that enables curation of models by subject matter experts. A method and interface will be developed to facilitate human-model interaction that enables formal definition of modeling problems and curation of automatically constructed models by users who are not data scientists.

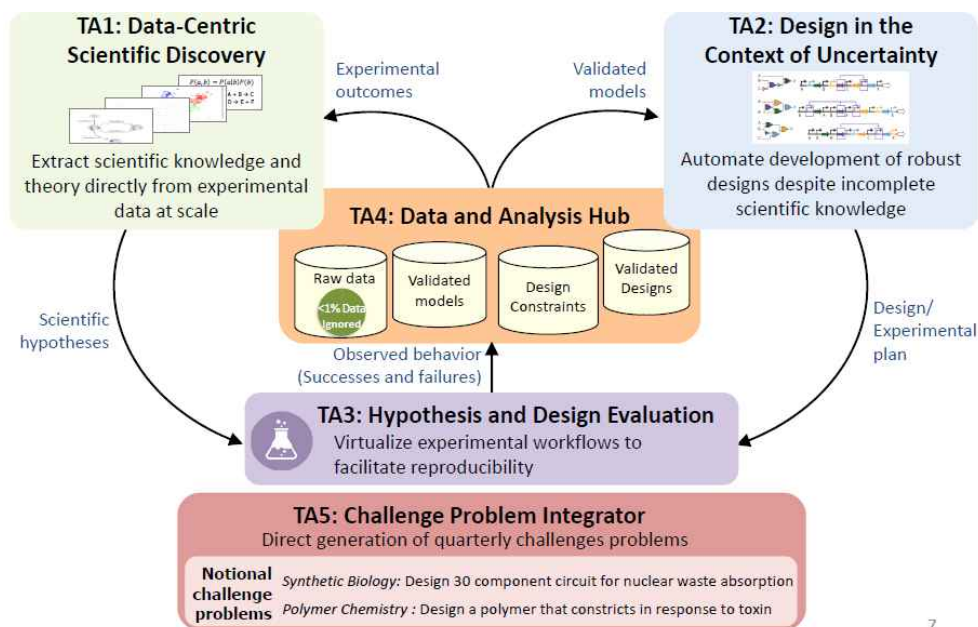
[DARPA 예비후보4] Synergistic Discovery and Design (SD2)

PM: Dr. Jennifer Roberts ※'16.11.22 공지

해당 분야에 아직 검증된 과학적 설명 모형이 존재하지 않음에도 불구하고 대용량 데이터를 이용해서 견고하게 실험 설계를 할 수 있는 방법을 고안하는 것을 목표로 하고 있고, 이를 통해 과학 발견을 가속화하고자 함 (5가지 TA로 구성됨)



The Synergistic Discovery and Design (SD2) program aims to develop data-driven methods to accelerate scientific discovery and robust design in domains that lack complete models. Engineers regularly use high-fidelity simulations to create robust designs in complex domains such as aeronautics, automobiles, and integrated circuits. In contrast, robust design remains elusive in domains such as synthetic biology, neuro-computation, and polymer chemistry due to the lack of high-fidelity models. SD2 seeks to develop tools to enable robust design despite the lack of complete scientific models.



[DARPA 예비후보5] Probabilistic Programming for Advancing Machine Learning (PPAML)

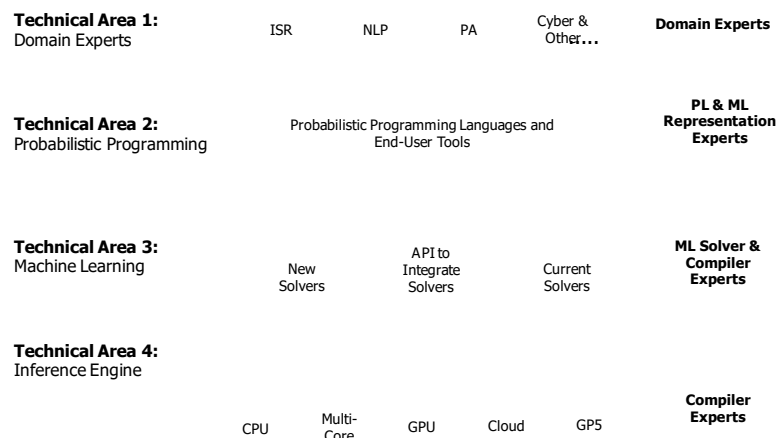
PM: Kathleen Fisher / Dr. Jennifer Roberts ※ '13.10.~'17.7 (46개월)

프로그래밍 스킬이 낮은 사용자도 쉽게 사용할 수 있는 확률적 프로그래밍 패러다임 기반의 기계학습 툴 개발 (프로그래밍 언어, 컴파일러, 재사용 가능 모델 라이브러리 포함)

<i>Traditional Programming</i>	<i>Probabilistic Programming</i>	
• Application	• Model	Code models capture how the data was generated using random variables to represent uncertainty
• Code Libraries	• Model Libraries	Libraries contain common model components: Markov chains, deep belief networks, etc.
• Programming Language	• Probabilistic Programming Language	PPL provides probabilistic primitives & traditional PL constructs so users can express model, queries, and data
• Compiler	• Inference Engine	Inference engine analyzes probabilistic program and chooses appropriate solver(s) for available hardware
• Hardware	• Hardware	Hardware can include multi-core, GPU, cloud-based resources, GraphLab, UPSIDE/Analog Logic results, etc.
High-level programming languages facilitate building complex systems Probabilistic programming languages facilitate building rich ML applications		

The Probabilistic Programming for Advancing Machine Learning (PPAML) program aims to address these challenges. Probabilistic programming is a new programming paradigm for managing uncertain information. Using probabilistic programming languages, PPAML seeks to greatly increase the number of people who can successfully build machine learning applications and make machine learning experts radically more effective. PPAML has five specific tactical objectives:

- Shorten machine learning model code to make models faster to write and easier to understand
- Reduce development time and cost to encourage experimentation
- Facilitate the construction of more sophisticated models that incorporate rich domain knowledge and separate queries from underlying code
- Reduce the level of expertise necessary to build machine learning applications
- Support the construction of integrated models across a wide variety of domains and tool types



주) Surveillance and Reconnaissance (ISR), Natural Language Processing (NLP), Predictive Analytics (PA)

[DARPA 예비후보6] Mining and Understanding Software Enclaves (MUSE)

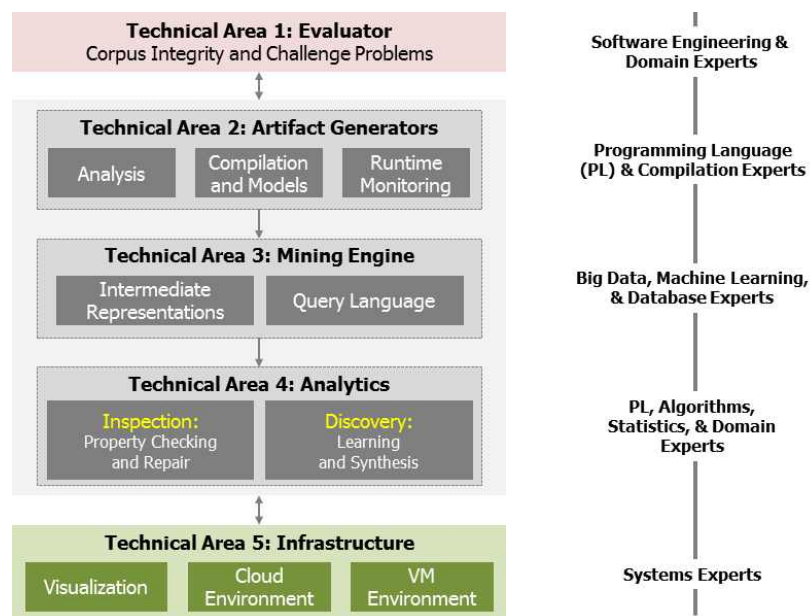
PM: Dr. Sandeep Neema ※'14.8~18.11 (52개월)

※ Rice University의 PLINY 프로젝트 (1,100만 달러 투자)

SW 품질 제고를 위한 대용량의 오픈소스 코드를 활용한 빅데이터 분석 톨 개발

Testing	Static Program Analysis	Dynamic Program Analysis
<p><i>PLT Redex</i>: DSL for specifying, debugging, and testing operational semantics</p> <p><i>Quickcheck</i>: Specification-driven formulation of properties that can be checked using random testing</p> <p><i>Csmith</i>: Random generator of C programs that conform to C99 standard for stress-testing compilers, analyses, etc.</p> <p><i>CUTE</i>: Unit-testing of C programs with pointer arguments by combining symbolic and concrete executions</p> <p><i>Korat</i>: Constraint-based generation of complex test inputs for Java programs, focusing on data structures and invariants</p>	<p><i>CFA</i>: Whole-program control-flow analysis that computes the set of procedures that can be invoked at a call-site</p> <p><i>ASTREE</i>: Abstract interpretation of real-time embedded software designed to prove absence of runtime errors by overapproximation of program behavior</p> <p><i>TVLA</i>: Flow-sensitive shape analysis of dynamically-allocated imperative data structures</p> <p><i>Bddbddbt</i>: Context- and field-sensitive analysis applied to Java that translates analysis rules expressed in Datalog to BDD representation</p> <p><i>Saturn</i>: Scalable and modular summary-driven bit-level constraint-based analysis framework</p> <p><i>Coverity</i>: Unsound scalable analyses used to check correctness of C, C++, and Java programs.</p>	<p><i>Contracts</i>: Assertions checked at runtime with blame</p> <p><i>Daikon</i>: Likely pre- and post-condition invariant detection over propositional terms, based on program instrumentation and</p> <p><i>Valgrind</i>: Instrument binary programs to track memory access violations and data races using dynamic recompilation</p> <p><i>Fastrack</i>: Lightweight data race detector that uses vector clocks and a dynamically constructed happens-before relation</p>
Symbolic Execution	Logics and Types	Model Checking
<p><i>KLEE</i>: Symbolic execution engine to generate high-coverage test cases</p> <p><i>S2E</i>: Scalable path-sensitive platform</p> <p><i>Jstor</i>, <i>Space Invader</i>, <i>Smallfoot</i>: Separation-logic based tools for verifying expressive shape properties of dynamic data structures and heaps</p> <p><i>ESC</i>: Extended static checking that combines type checking with theorem proving</p>	<p><i>Coq</i>, <i>Agda</i>, <i>Isabelle</i>, <i>ACL2</i>, <i>NuPRL</i>: Mechanized proof assistants</p> <p><i>Ynot</i>: Hoare Type Theory</p> <p><i>Rely-Guarantee Reasoning</i>: Modular verification of shared-memory concurrency</p>	<p><i>CVC</i>, <i>SLAM</i>, <i>Blast</i>, <i>Spin</i>, <i>Java PathFinder</i></p> <p><i>CHES</i>: Bounded model-checking for unit-testing of shared-memory concurrent programs</p> <p><i>TLA</i>: Temporal logic of actions for specifying and checking concurrent systems</p> <p><i>Liquid Type Inference</i>: Discovery of expressive refinement properties in Haskell, ML, and C</p> <p><i>Hybrid Type Checking and Soft Typing</i></p> <p><i>Session Types</i>: type systems for expressing communication protocols</p>

MUSE seeks to make significant advances in the way software is built, debugged, verified, maintained and understood. Central to its approach is the creation of a community infrastructure built around a large, diverse and evolving corpus of software drawn from the hundreds of billions of lines of open source code available today.



[DARPA 예비후보7] SAFEWARE

PM: Dr. Carey Schwartz ※ '14.9월 공지, 총 48개월 (12월짜리 4 단계)

SW 코드 역공학을 막는 솔루션 개발

DARPA's SafeWare program aims to develop obfuscation technology that would render the intellectual property in software (e.g., proprietary algorithms) incomprehensible to a reverse engineer, but allow the code to otherwise compile and run normally. To accomplish this, SafeWare researchers aim to develop fundamentally new program obfuscation technology with (i) quantifiable security that (ii) depends not on the appearance of complexity in code structure, but on the difficulty of the mathematical problems an attacker would have to solve to successfully de-obfuscate the program.

To gain the security benefits of program obfuscation, a price in program runtime efficiency must be paid. Fortunately, recent developments indicate that the scaling between the price paid in efficiency and the security benefit gained by obfuscation is favorable. Extant theory guarantees an adversary work factor (i.e., CPU cycles required to break the obfuscation) that scales exponentially with respect to polynomial increases in program runtime. Unfortunately, this runtime overhead is still extremely large in absolute terms, making even the simplest kinds of programs run unacceptably slow. SafeWare will address the main practical obstacle to implementing this technology today: reducing these overheads so that software can run efficiently for users while being safe from reverse engineering.

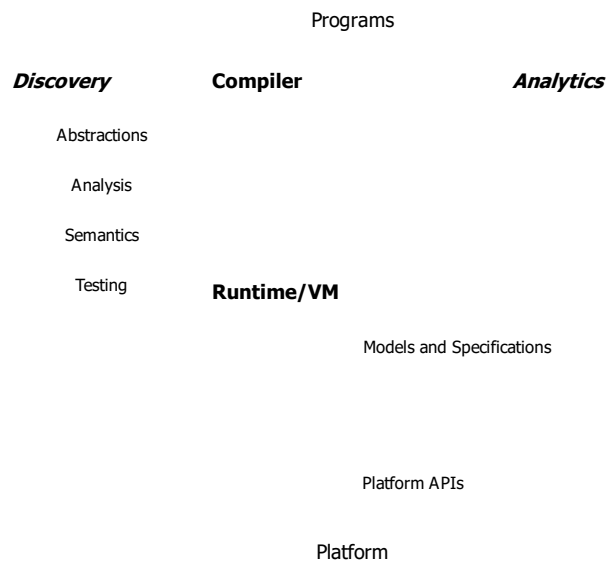
- **TA1: Mathematical Foundations of Program Obfuscation**
- **TA2: Implementation of Program Obfuscation**
- **TA3: Proof-of-Principle Demonstrations**
- **TA4: Testing, Evaluation and Research Integration**

If successful, SafeWare technologies will provide provably-secure protection of sensitive intellectual property and algorithmic information in software that is vulnerable to capture and dissection.

[DARPA 예비후보8] Building Resource Adaptive Software Systems (BRASS)

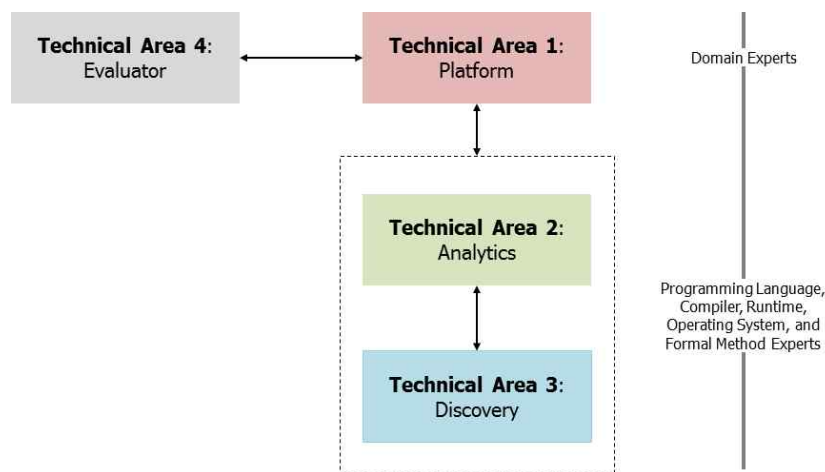
PM: Dr. Sandeep Neema ※ '15.10~19.9 (48개월/3단계)

The goal of the Building Resource Adaptive Software Systems program (BRASS) is to realize foundational advances in **the design and implementation of long-lived, survivable and complex software systems that are robust to changes in the physical and logical resources provided by their ecosystem**. These advances will necessitate integration of new resource-aware program abstractions and analyses, in addition to novel compiler and systems designs to trigger adaptive transformations and validate their effectiveness.



The final products of this program are as follow:

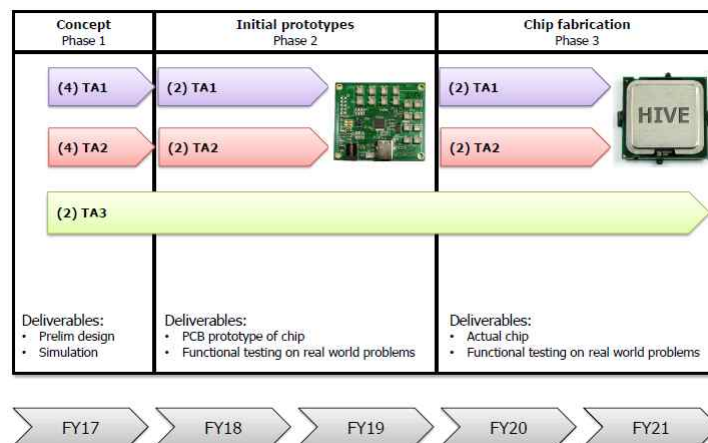
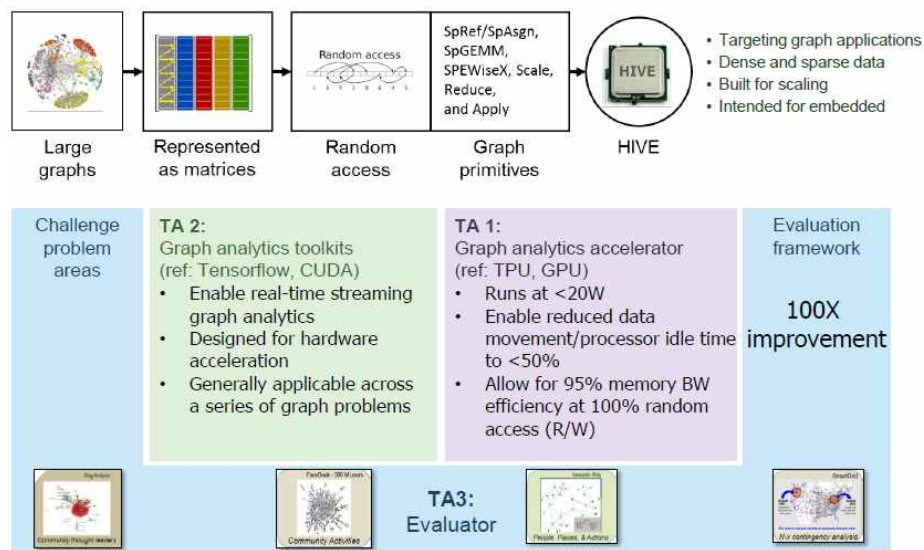
- Fully adaptive platforms
- Adaptive monitoring and transformation tools
- Continuous resource adaptive analysis suite



[DARPA 예비후보9] Hierarchical Identify Verify Exploit (HIVE)

PM: Mr. Wade Shen ※ '17~21년

The HIVE program is looking to build a graph analytics processor that can process streaming graphs 1000X faster and at much lower power than current processing technology. If successful, the program will enable graph analytics techniques powerful enough to solve tough challenges in cyber security, infrastructure monitoring and other areas of national interest. Graph analytic processing that currently requires racks of servers could become practical in tactical situations to support front-line decision making. What 's more, these advanced graph analytics servers could have the power to analyze the billion- and trillion-edge graphs that will be generated by the Internet of Things, ever-expanding social networks, and future sensor networks.



In parallel with the hardware development of a HIVE processor, DARPA is working with MIT Lincoln Laboratory and Amazon Web Services (AWS) to host the **HIVE Graph Challenge** with the goal of developing a trillion-edge dataset. This freely available dataset will spur innovative software and hardware solutions in the broader graph analysis community that will contribute to the HIVE program.

<부록 11> IARPA 예비 후보

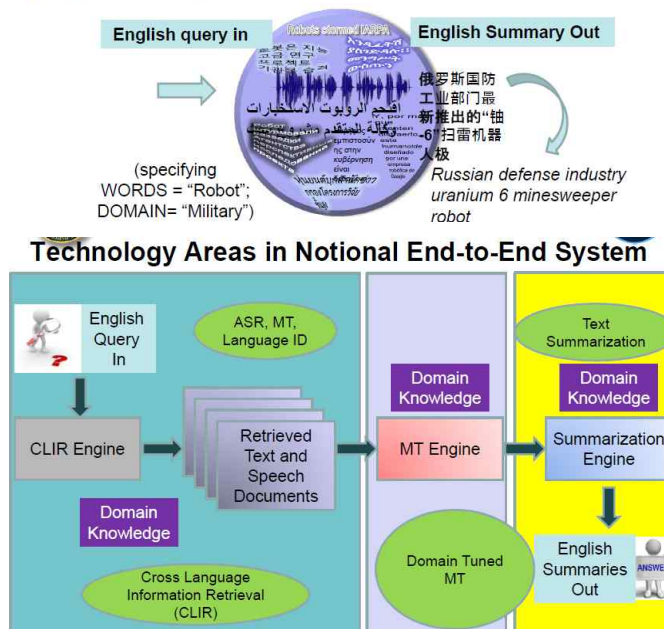
[IARPA 예비후보1] Machine Translation for English Retrieval of Information in Any Language (MATERIAL)

The MATERIAL Program seeks to develop methods for finding speech and text content in **low-resource languages** that is relevant to domain-contextualized English queries. Such methods must use **minimal training data** and be rapidly deployable to new languages and domains. Content that is responsive to the queries will be returned from multiple genres along with succinct summaries in English. The overall end-to-end capability will enable monolingual triage of multilingual datasets.

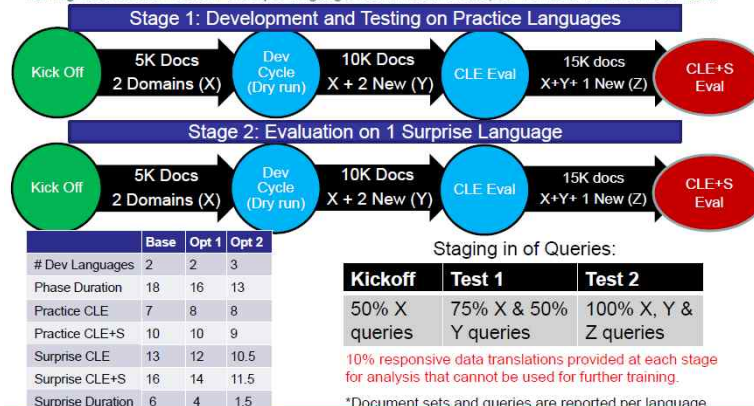
PM: Carl Rubino ※ '17.3.20 제안서 접수 마감, 총 47개월 수행

영어로 질문을 하면 다국어 자료를 검색한 후 검색 결과를 다시 영어로 요약하여 제시 (단, 도메인 특화된 검색어 사용, 상대적으로 적은 학습 데이터 이용)

- An “English-in, English-out” information retrieval system that, given a domain-sensitive English query, will retrieve relevant data from a large multilingual repository and display the retrieved information in English as summaries that reflect the document relevance:



Training Data at Each Kickoff Period per language: 800K Words Bitexts; 35-45 Hours of Transcribed Audio



[IARPA 예비후보2] Deep Intermodal Video Analytics (DIVA): 멀티 카메라 동영상 속 동작 인식 및 이해

PM: Terry Adams ※ '17.7~'21.6 (48개월)

복수의 스트리밍 카메라 환경에서의 자동 객체 및 행위 인식 방법 개발

DIVA program seeks to develop **robust automatic activity detection** for a **multi-camera streaming video environment**. Activities will be enriched by **person and object detection**. DIVA will address activity detection for both **forensic applications** and for **real-time alerting**.

- **Two main goals are evaluated:**
 - (a) **Activity detection** - The DIVA program intends to develop robust automated activity detection for a multi-camera streaming video environment.

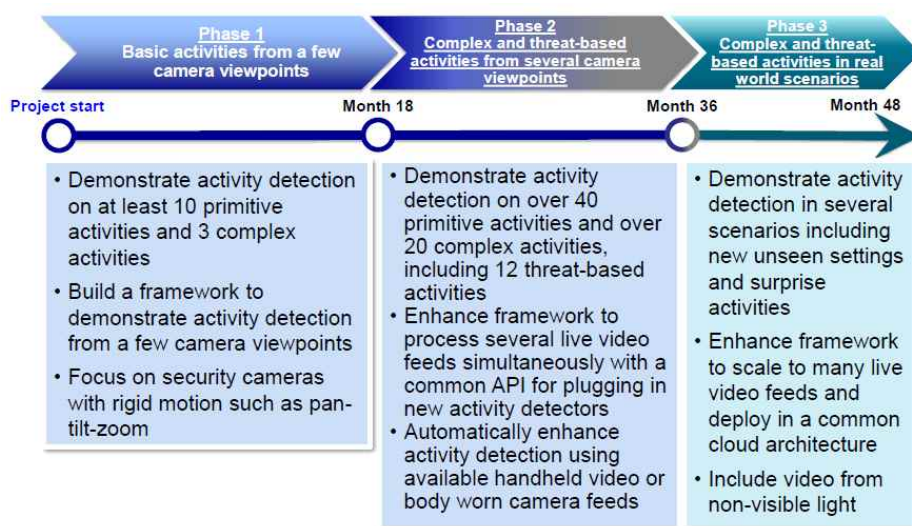


- (b) **Person/object detection** - DIVA activities will be enriched by person and object detection, as well as recognition at multiple levels of granularity.



DIVA is anticipated to be a three-phase program. The program will focus on three major thrusts throughout all phases:

- Detection of **primitive activities** occurring in ground-based video collection
- Detection of **complex activities**, including pre-specified or newly defined activities
- Person and object detection and recognition across multiple overlapping and non-overlapping camera viewpoints

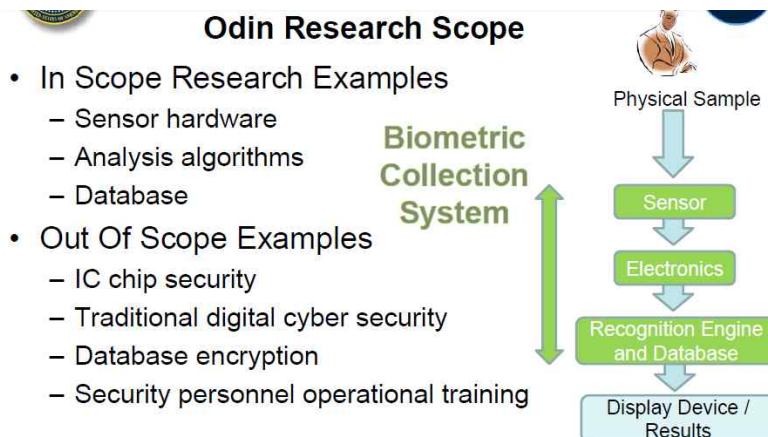
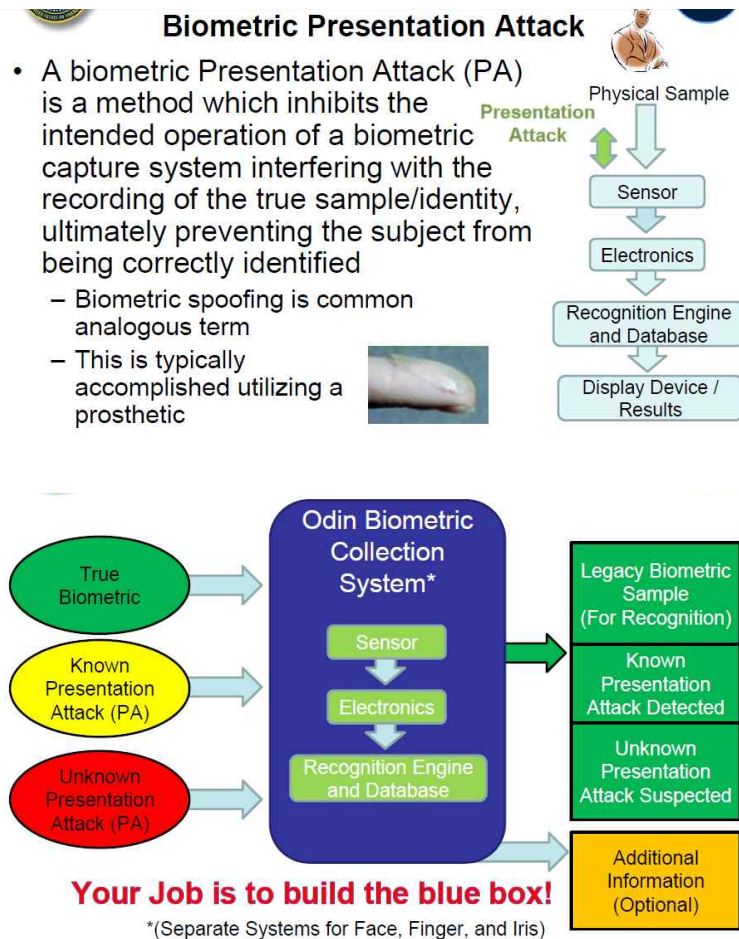


[IARPA 예비후보3] Odin

PM: Chris Boehnen ※ '17.3~'21.2 (48개월/3단계) ※ SRI International 수행중 (1,250만 달러)

알려진 또는 새로운 생체정보 도용 수법 탐지 방법 개발

The goal of this program is to utilize Presentation Attack Detection (PAD) to identify known and unknown Presentation Attacks (PA) in a biometric collection system.



초전도 방식의 슈퍼컴퓨터 개발을 위한 초전도 전자기기 설계 툴 개발

Program Objectives

SuperTools seeks to

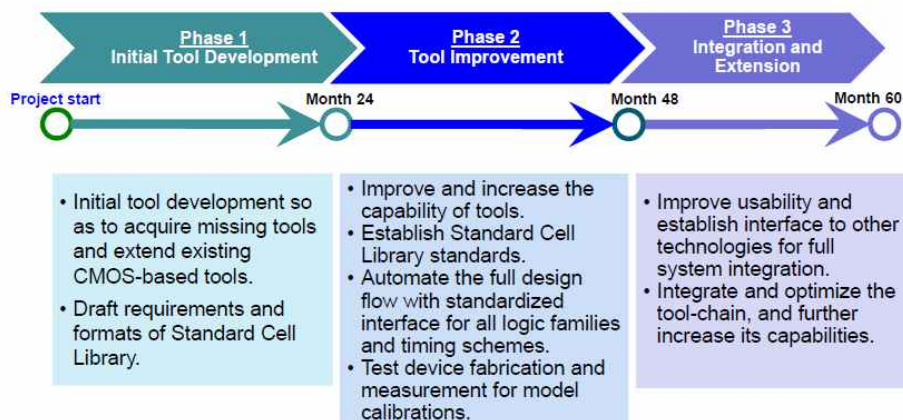
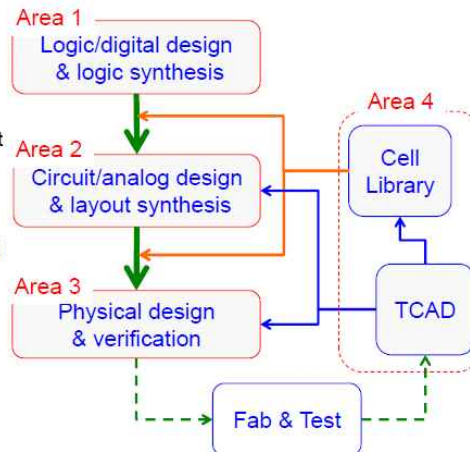
- Develop a comprehensive set of **Electronic Design Automation (EDA)** tools to enable **Very-Large-Scale Integration (VLSI)** design of **Superconducting Electronics (SCE)** from HDL to GDSII,
- Develop physics-based **Technology CAD (TCAD)** tools to enable device and process simulations and device parameter extractions for better design-to-hardware fidelity, and
- Establish **open, interoperable Standard Cell Library formats** as standardized interface to bridge collaboration between foundry and designers and to speed up technology development.

Target SCE design capability:

- Phase I: support design of circuits with **>100k JJs** (or 10k gates)
- Phase II: support design of circuits with **>1M JJs** (or 100k gates)
- Phase III: support design of circuits with **>10M JJs** (or 1M gates)

SuperTools Approach

- To concurrently focus on Four Program Focus Areas, addressing:
 - Each major step of the EDA design flow, and
 - Back-end support with development of Cell Library standards & TCAD tools.
- To incorporate SCE-Specific design considerations and features into tool development.



<부록12> NSF CPS 예비 후보

[NSF CPS 예비후보1]

Building a Smart City Economy and Information Ecosystem to Motivate Pro-Social Transportation Behavior

CPS: TTP Option: Medium:

Award Number:1739413; Principal Investigator:Alexandros Labrinidis; Co-Principal Investigator:Adam Lee, Sera Linardi, Mark Magalotti, Konstantinos Pelechrinis; Organization:University of Pittsburgh;NSF Organization:CNS Start Date:09/01/2017; Award Amount:\$1,439,999.00; ~August 31, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1739413&HistoricalAwards=false

The growth and expansion of cities since the mid 20th century has led to a strong dependency on private automobiles. During the last years, urban planners have started rethinking the mobility modes in a city and have finally realized that a truly sustainable transportation and urban environment in general, requires a shift to multimodal transportation. In the PittSmartLiving project, we view the shift to multimodal transportation in a holistic way. In particular, we will design, develop, deploy, and evaluate a platform that will integrate information from and align the incentives of all involved stakeholders (commuters, transport operators, and local businesses) towards increasing the utilization and quality of public transportation. The resulting Cyber-Physical system will (1) provide commuters with **real-time information** of arrival and utilization of all relevant options of public transit (e.g., bus, subway, shuttles, bikes, etc.), and (2) build a **marketplace** around multimodal mobility, where businesses can offer time-sensitive incentives connected to this transit information to nearby commuters (e.g., the next bus is too full, come in and enjoy \$1 coffee). This has the potential to improve not only the overall ridership experience by balancing utilization across public transportation networks (e.g., shifting some of the demand away from the peak hours), but also to optimize customer flows in local businesses. Significant emphasis will be given to the development of mechanisms that will be able to deliver the required services while respecting the privacy expectations of the commuters. As part of this project, an unprecedented experimental infrastructure will be deployed (in the **Oakland and Downtown areas of Pittsburgh**) that will allow the PIs to identify a set of incentive mechanisms that can shift commuters to public transportation in a real urban environment. This is the first time that an urban core truly becomes a laboratory, where scientists and engineers can run experiments aimed at improving the quality of life of city-dwellers.

The main expected technical contributions of this project can be summarized as follows. (1) Development of a holistic urban transportation system that balances utilization across both public transportation networks and local businesses, thus improving not only public transit but also general urban living. (2) Design and evaluation of the market mechanism that integrates and aligns the incentives of various stakeholders. (3) Shift of attention from temporal efficiency (i.e., fastest route) to more sustainable commuting (e.g., public transit, biking etc.) as well as commuting options geared towards the well-being of dwellers (e.g., "beautiful" routes, "clean" routes, "accessible" routes etc.)

Additional information about this project can be found at <http://PittSmartLiving.org>

Partners

TransitScreen
Port Authority of Allegheny County
Healthy Ride
City of Pittsburgh
Oakland Business Improvement District
Pittsburgh Downtown Partnership
Envision Downtown
Oakland Transportation Management Association
Pittsburgh 2030 District
Radius Networks
UPMC
University of Pittsburgh:
Department of Parking, Transportation & Services
Office of Community and Governmental Relations
Center for Social & Urban Research

[NSF CPS 예비후보2]

Security and Privacy-Aware Cyber-Physical Systems

Synergy: Collaborative:

Award Number:1505799; Principal Investigator:Insup Lee; Co-Principal Investigator:Oleg Sokolsky, George Pappas, Andreas Haeberlen, Nadia Heninger; Organization:University of Pennsylvania;NSF Organization:CNS Start Date:09/01/2015; Award Amount:\$**1,125,000.00**; ~August 31, 2018 (Estimated)

Security and privacy concerns in the increasingly interconnected world are receiving much attention from the research community, policymakers, and general public. However, much of the recent and on-going efforts concentrate on security of general-purpose computation and on privacy in communication and social interactions. The advent of cyber-physical systems (e.g., safety-critical IoT), which aim at tight integration between distributed computational intelligence, communication networks, physical world, and human actors, opens new horizons for intelligent systems with advanced capabilities. These systems may reduce number of accidents and increase throughput of transportation networks, improve patient safety, mitigate caregiver errors, enable personalized treatments, and allow older adults to age in their places. At the same time, cyber-physical systems introduce new challenges and concerns about safety, security, and privacy. The proposed project will lead to **safer, more secure and privacy preserving CPS**. As our lives depend more and more on these systems, specifically in **automotive, medical, and Internet-of-Things domains**, results obtained in this project will have a direct impact on the society at large. The study of emerging legal and ethical aspects of large-scale CPS deployments will inform future policy decision-making. The educational and outreach aspects of this project will help us build a workforce that is better prepared to address the security and privacy needs of the ever-more connected and technologically oriented society.

Cyber-physical systems (CPS) involve tight integration of computational nodes, connected by one or more communication networks, the physical environment of these nodes, and human users of the system, who interact with both the computational part of the system and the physical environment. Attacks on a CPS system may affect all of its components: computational nodes and communication networks are subject to malicious intrusions, and physical environment may be maliciously altered. CPS-specific security challenges arise from two perspectives. On the one hand, conventional information security approaches can be used to prevent intrusions, but attackers can still affect the system via the physical environment. **Resource constraints**, inherent in many CPS domains, may prevent heavy-duty security approaches from being deployed. This proposal will develop a framework in which the mix of prevention, detection and recovery, and robust techniques work together to improve the security and privacy of CPS. Specific research products will include techniques providing: 1) accountability-based detection and bounded-time recovery from malicious attacks to CPS, complemented by novel preventive techniques based on lightweight cryptography; 2) security-aware control design based on attack resilient state estimator and sensor fusions; 3) privacy of data collected and used by CPS based on differential privacy; and, 4) evidence-based framework for CPS security and privacy assurance, taking into account the operating context of the system and human factors. Case studies will be performed in applications with **autonomous features of vehicles, internal and external vehicle networks, medical device interoperability, and smart connected medical home**.

J. Park, M. Pajic, I. Lee, and O. Sokolsky. "Scalable Verification of Linear Controller Software," *22nd International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, 2016.

Hung Nguyen, Bipeen Acharya, Radoslav Ivanov, Andreas Haeberlen, Linh T.X. Phan, Oleg Sokolsky, Jesse Walker, James Weimer, William Hanson, and Insup Lee. "Cloud-based Secure Logger For Medical Devices," *1st International Workshop on Security, Privacy, and Trustworthiness in Medical Cyber-Physical Systems (MedSPT'16)*, 2016. 등 30편

[NSF CPS 예비후보3]

Secure Perception for Autonomous Systems

CAREER:

Award Number:1454474; Principal Investigator:Todd Humphreys; Co-Principal Investigator;; Organization:University of Texas at Austin;NSF Organization:ECCS Start Date:04/01/2015; Award Amount:\$500,000.00; ~March 31, 2020 (Estimated)

Today's automobiles are increasingly autonomous. The latest Mercedes S-class sedan applies corrective action when its driver strays out of lane or tailgates too closely. Semi-autonomy will soon yield to full autonomy. Nissan has promised a line of self-driving cars by 2020. Maritime craft are likewise moving from rudimentary autopilots to full autonomy, and autonomous aerial vehicles will doubtless play a significant role in the future economy.

Current versions of these vehicles are cocooned in an array of sensors, but neither the sensors nor the timing, navigation, and collision avoidance algorithms they feed have been designed for security against malicious attacks. Radar and acoustic sensors transmit predictable, uncoded signals; vehicle-to-vehicle communication protocols are either unauthenticated or critically dependent on insecure civil GPS signals (or both); and vehicle state estimators are designed for robustness but not security. These vulnerabilities are not merely conceptual: GPS spoofing attacks have been demonstrated against a drone and an ocean vessel, causing the drone to crash and the vessel to veer off course; likewise, it appears possible to cause road accidents by fooling a car's radar sensor into thinking a crash is imminent, thus triggering automatic braking.

This proposal seeks funding to fix these vulnerabilities by developing sensors and high-level decision-making algorithms that are hardened against such so-called field attacks. The goal of **secure control systems** is to survive and operate safely despite sensor measurements or control commands being compromised. This proposal focuses on an emergent category of cyber-physical attack that has seen little scrutiny in the secure control literature. Like cyber attacks, these attacks are hard to detect and can be executed from a distance, but unlike cyber attacks, they are effective even against control systems whose software, data, and communications networks are secure, and so can be considered a more menacing long-term threat. These are attacks on the physical fields such as electromagnetic, magnetic, acoustic, etc. measured by system sensors. As specialized sensor attacks, field attacks seek to compromise a system's perception of reality non-invasively from without, not from within. We emphasize field attacks against navigation, collision avoidance, and synchronization sensors, as these are of special importance to the rise of autonomous vehicles and the smart grid. This proposal's goal is to develop a coherent analytical foundation for secure perception in the presence of field attacks and to develop a suite of algorithms and tools to detect such attacks. A key insight behind this proposal's approach is that the physics of field attacks impose fundamental difficulties on the attacker that can be exploited and magnified to enable attack detection. This work will progressively build security into navigation, collision avoidance, and timing perception from the physical sensory layer to the top-level state estimation algorithms. The outcome of this work will be smarter, more skeptical sensor systems for autonomous vehicles and other autonomous systems.

Narula, Lakshay and Humphreys, Todd E. "Requirements for Secure Wireless Time Transfer," Proceedings of the IEEE/ION PLANS Meeting, 2016.

Mark L. Psiaki and Todd E. Humphreys. "{GNSS} spoofing and detection," Proceedings of the IEEE, 2016.

Lakshay Narula and Todd Humphreys. "Requirements for Secure Wireless Time Transfer," Proceedings of the IEEE/ION PLANS Conference, 2016.

관련 자료: Perception, Planning, Control, and Coordination for Autonomous Vehicles (Scott Drew Pendleton et al., 17 February 2017, MDPI)

<https://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjVp56266vXAhULfLwKHeDuBi0QFggtMAA&url=http%3A%2F%2Fwww.mdpi.com%2F2075-1702%2F5%2F1%2F6%2Fpdf-vor&usq=AOvVaw2G8Quq8MnNYegmh7UG2ttK>

[NSF CPS 예비후보4]

A Cyber-Physical System for Monitoring, Analysis and Mitigation of Urban Noise Pollution

CPS: Frontier: SONYC:

Award Number:1544753; Principal Investigator:Juan Bello; Co-Principal Investigator:Anish Arora, Claudio Silva, Oded Nov, Roger DuBois; Organization:New York University;NSF Organization:CNS Start
Date:08/01/2016; Award Amount:\$**2,015,458.00**; ~July 31, 2021 (Estimated)

This Frontier award supports the SONYC project, a smart cities initiative focused on developing a **cyber-physical system (CPS) for the monitoring, analysis and mitigation of urban noise pollution**. Noise pollution is one of the topmost quality of life issues for urban residents in the U.S. with proven effects on health, education, the economy, and the environment. Yet, most cities lack the resources for continuously monitoring noise and understanding the contribution of individual sources, the tools to analyze patterns of noise pollution at city-scale, and the means to empower city agencies to take effective, data-driven action for noise mitigation. The SONYC project advances novel technological and socio-technical solutions that help address these needs.

SONYC includes a distributed network of both sensors and people for large-scale noise monitoring. The sensors use low-cost, low-power technology, and cutting-edge machine listening techniques, to produce calibrated acoustic measurements and recognizing individual sound sources in real time. Citizen science methods are used to help urban residents connect to city agencies and each other, understand their noise footprint, and facilitate reporting and self-regulation. Crucially, SONYC utilizes big data solutions to analyze, retrieve and visualize information from sensors and citizens, creating a comprehensive acoustic model of the city that can be used to identify significant patterns of noise pollution. This data can in turn be used to drive the strategic application of noise code enforcement by city agencies, in a way that optimally reduces noise pollution. The entire system, integrating cyber, physical and social infrastructure, forms a closed loop of continuous sensing, analysis and actuation on the environment.

SONYC is an interdisciplinary collaboration between researchers at New York University and Ohio State University. It provides multiple educational opportunities to students at all levels, including an outreach initiative for K-12 STEM education. The project uses New York City as its focal point, involving partnerships with the city's Department of Environmental Protection, Department of Health and Mental Hygiene, the business improvement district of Lower Manhattan, and ARUP, one of the world's leaders in environmental acoustics. SONYC is an innovative and high-impact application of cyber-physical systems to the realm of smart cities, and potentially a catalyst for new CPS research at the intersection of engineering, data science and the social sciences. It provides a blueprint for the mitigation of noise pollution that can be applied to cities in the US and abroad, potentially affecting the quality of life of millions of people.

Mydlarz, C., Salamon, J., and Bello, J.P.. "The Implementation of Low-cost Urban Acoustic Monitoring Devices," Applied Acoustics, Special Issue on Acoustics for Smart Cities, v.117, 2017, p. 207.

Salamon, J., and Bello, J.P.. "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification," IEEE Signal Processing Letters, v.24, 2017, p. 279.

[NSF CPS 예비후보5]

Foundations of Resilient CybEr-Physical Systems (FORCES)

CPS: Frontiers: Collaborative Research:

This NSF Cyber-Physical Systems (CPS) Frontiers project "Foundations Of Resilient CybEr-physical Systems (FORCES)" focuses on the resilient design of large-scale networked CPS systems that directly interface with humans. FORCES aims to provide **comprehensive tools** that allow the CPS designers and operators to combine **resilient control (RC) algorithms** with **economic incentive (EI) schemes**.

Scientific Contributions

The project is developing **RC tools** to withstand a wide-range of attacks and faults; learning and control algorithms which integrate human actions with spatio-temporal and hybrid dynamics of networked CPS systems; and model-based design to assure semantically consistent representations across all branches of the project. Operations of networked CPS systems naturally depend on the systemic social institutions and the individual deployment choices of the humans who use and operate them. The presence of incomplete and asymmetric information among these actors leads to a gap between the individually and socially optimal equilibrium resiliency levels. The project is developing EI schemes to reduce this gap. The core contributions of the FORCES team, which includes experts in control systems, game theory, and mechanism design, are the foundations for the co-design of RC and EI schemes and technological tools for implementing them.

Expected Impacts

Resilient CPS infrastructure is a critical National Asset. FORCES is contributing to the development of new Science of CPS by being the first project that integrates networked control with game theoretic tools and the economic incentives of human decision makers for resilient CPS design and operation. The FORCES integrated co-design philosophy is being validated on two CPS domains: electric power distribution and consumption, and transportation networks. These design prototypes are being tested in real world scenarios. The team's research efforts are being complemented by educational offerings on resilient CPS targeted to a large and diverse audience.

Award Number:1239054; Principal Investigator:Saurabh Amin; Co-Principal Investigator:Asuman Ozdaglar, Hamsa Balakrishnan; Organization:Massachusetts Institute of Technology;NSF Organization:CNS Start Date:04/15/2013; Award Amount:**\$2,150,000.00**; ~March 31, 2018 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1239054&HistoricalAwards=false

Award Number:1239166; Principal Investigator:S. Shankar Sastry; Co-Principal Investigator:Dawn Song, Alexandre Bayen, Claire Tomlin; Organization:University of California-Berkeley;NSF Organization:CNS Start Date:04/15/2013; Award Amount:**\$3,600,000.00**; ~March 31, 2018 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1239166&HistoricalAwards=false

Award Number:1238959; Principal Investigator:Xenofon Koutsoukos; Co-Principal Investigator:Gabor Karsai, Janos Sztipanovits; Organization:Vanderbilt University;NSF Organization:CNS Start Date:04/15/2013; Award Amount:**\$1,945,299.00**; ~March 31, 2018 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1238959&HistoricalAwards=false

Award Number:1238962; Principal Investigator:Demosthenis Teneketzis; Co-Principal Investigator:Ian Hiskens; Organization:University of Michigan Ann Arbor;NSF Organization:CNS Start Date:04/15/2013; Award Amount:**\$1,300,000.00**; ~March 31, 2018 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1238962&HistoricalAwards=false

[NSF CPS 예비후보6]

Correct-by-Design Control Software Synthesis for Highly Dynamic Systems

CPS: Frontier: Collaborative Research:

This CPS Frontiers project addresses highly dynamic Cyber-Physical Systems (CPSs), understood as systems where a computing delay of a few milliseconds or an incorrectly computed response to a disturbance can lead to catastrophic consequences. Such is the case of cars losing traction when cornering at high speed, unmanned air vehicles performing critical maneuvers such as landing, or disaster and rescue response bipedal robots rushing through the rubble to collect information or save human lives. The preceding examples currently share a common element: the design of their control software is made possible by extensive experience, laborious testing and fine tuning of parameters, and yet, the resulting closed-loop system has no formal guarantees of meeting specifications.

The vision of the project is **to provide a methodology that allows for complex and dynamic CPSs to meet real-world requirements in an efficient and robust way through the formal synthesis of control software**. The research is developing a formal framework for correct-by-construction control software synthesis for highly dynamic CPSs with broad applications to automotive safety systems, prostheses, exoskeletons, aerospace systems, manufacturing, and legged robotics.

The design methodology developed here will improve the competitiveness of segments of industry that require a tight integration between hardware and highly advanced control software such as: **automotive (dynamic stability and control)**, aerospace (UAVs), medical (prosthetics, orthotics, and exoskeleton design) and robotics (legged locomotion). To enhance the impact of these efforts, the PIs are developing interdisciplinary teaching materials to be made freely available and disseminating their work to a broad audience.

Award Number:1239085; Principal Investigator:Paulo Tabuada; Co-Principal Investigator;; Organization:University of California-Los Angeles;NSF Organization:CNS Start Date:04/01/2013; Award Amount:\$**1,000,000.00**; ~March 31, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1239085&HistoricalAwards=false

A Mehra, WL Ma, F Berg, P Tabuada, JW Grizzle, AD Ames. "Adaptive cruise control: Experimental validation of advanced controllers on scale-model cars," American Control Conference (ACC), 2015. 등 38편

Award Number:1239037; Principal Investigator:Jessy Grizzle; Co-Principal Investigator:Kevin Galloway; Organization:University of Michigan Ann Arbor;NSF Organization:CNS Start Date:04/01/2013; Award Amount:\$**1,600,000.00**; ~March 31, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1239037&HistoricalAwards=false

R. Postoyan, P. Tabuada, D. Nesic, A. Anta. "A Framework for the Event-Triggered Stabilization of Nonlinear Systems," IEEE Transactions on Automatic Control, v.60, 2015, p. 982. 등 3편

<부록 13> NSF 빅데이터 예비 후보

[NSF 빅데이터 예비후보01]

A Formal Foundation for Big Data Management

BIGDATA: Mid-Scale: DCM:

Award Number:1247469; Principal Investigator:Dan Suciu; Co-Principal Investigator:Magdalena Balazinska, Bill Howe; Organization:University of Washington;NSF Organization:IIS Start Date:01/01/2013; Award Amount:\$2,966,667.00; ~December 31, 2017 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1247469&HistoricalAwards=false

The ability to analyze massive-scale datasets has become an important tool both in industry and in the sciences and many systems have recently emerged to support it. However, effective methods for deep data analytics are currently high-touch processes: they require **a highly specialized expert** who thoroughly understands the application domain and pertinent disparate data sources and who needs to perform repeatedly a series of data exploration, manipulation and transformation steps to prepare the data for querying, machine learning or data mining algorithms. This project explores the foundations of big data management with the ultimate goal of significantly **improving the productivity in big data analytics** by accelerating **the bottleneck step of data exploration**. The project integrates two thrusts: a theoretical study, which leads to new fundamental results regarding the complexity of various new (ad hoc) data transformations in modern massive-scale systems, and a systems study, which leads to a multi-platform software middleware for expressing and optimizing ad hoc data analytics techniques. The middleware is designed to augment and integrate existing analytics solutions in order to facilitate and improve methods of interest to the community and compatible with many existing platforms.

The results of this project will make it easier for domain experts to conduct complex data analysis on big data and on large computer clusters. All research results will be released in a middleware package layered on top of existing big-data systems. The middleware includes all the new algorithms, optimization techniques, fault-tolerance and skew mitigation mechanisms, and generalized aggregates developed during the project. In addition, the project develops and deploys a **Web-based query-as-a-service interface to the new middleware**. The project Web site (<http://myriadb.cs.washington.edu>) provides access to the software, additional results and information. Project results will be included in educational and outreach activities in big data analytics, including new curricula at the undergraduate, graduate, and professional levels.

Jeremy Hyrkas, Sophie Clayton, Francois Ribalet, Daniel Halperin, E. Virginia Armbrust, and Bill Howe. "Cluster Analysis for Large-Scale, Continuous Flow Cytometry Data," *Bioinformatics*, 2015.

J Hyrkas, S Clayton, F Ribalet, D Halperin, EV Armbrust, B Howe. "Scalable clustering algorithms for continuous environmental flow cytometry," *Bioinformatics*, v.32, 2016, p. 417.

Brandon Haynes, Alvin Cheung, Magdalena Balazinska. "PipeGen: Data Pipe Generator for Hybrid Analytics," *Symposium on Cloud Computing*, 2016, p. 470.

Jingjing Wang, Magdalena Balazinska. "Toward elastic memory management for cloud data analytics.," *BeyondMR@SIGMOD*, 2016.

Parmita Mehta, Sven Dorkenwald, Dongfang Zhao, Tomer Kaftan, Alvin Cheung, Magdalena Balazinska, Ariel Rokem, Andrew J. Connolly, Jacob VanderPlas, Yusra AlSayyad. "Comparative Evaluation of Big-Data Systems on Scientific Image Analytics Workloads.," *Computing Research Repository*, 2016. [doi:10.1612.02485](https://doi.org/10.1612.02485) 등 10편

[NSF 빅데이터 예비후보02]

Discovery and Social Analytics for Large-Scale Scientific Literature

BIGDATA: Mid-Scale: ESCE: Collaborative Research:

Award Number:1247696; Principal Investigator:Rebecca Wright; Co-Principal Investigator:Paul Kantor; Organization:Rutgers University New Brunswick;NSF Organization:IIS Start Date:01/01/2013; Award Amount:\$1,004,784.00;~December 31, 2017 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1247696&HistoricalAwards=false

Big data analytics is, fundamentally, the problem of bringing the massive amounts of data produced today down to human scale. In particular scientists, engineers, physicians, and many others in knowledge-intensive professions face data that is beyond human scale. This data is in the repositories that collect the data and the reports or results in their fields. This project will address the problem of bringing all this knowledge under control by using even more data, namely the individual and social patterns of how these repositories are accessed and used, and user-specific judgments (valuations) of the data. The proposed research will develop novel **algorithms** and an **open-source infrastructure** for improving discovery within and access to data repositories. These algorithms will aggregate and analyze the social analytic data, gathered from professional communities of data users, and will motivate them to participate by providing recommendations.

The transformative goal is to develop methods for organizing, and operationalizing the access and preference patterns of users of large repositories, and for integrating those valuations to accelerate discovery within the collections. Diverse human minds interacting with data collections, as they carry out their own research or operational activities, provide a powerful source of information about the value of the data itself. Those data items may be textual documents, numerical datasets, or other kinds of media content. The novel methods for representing, aggregating, organizing and valuating interactions between the users and the items can reveal structures within data collections, which were previously invisible to any individual. This discovery of interrelations within data, driven by the capture of human intelligence, will accelerate the processes of scientific discovery. Users who are permitted to value data, and who are motivated by receiving valuable recommendations in return, reveal more about their own interests. This makes it possible to discover relations among the data items and among the users themselves. The educational goals are to: (a) contribute to the education of specific graduate students supported by the project, and undergraduates via the REU mechanism; (b) generate new educational materials related to algorithmic innovations, and to research findings; and (c) improve access to and discovery within specific collections of materials. Research findings will be included in courses at all three collaborating universities.

Additional information about the project (including publication, software, data sets) will be made available through the project web site: http://arxiv_xs.rutgers.edu/.

C Wang, D Blei. "Variational inference in nonconjugate models," *Journal of Machine Learning Research*, v.NA, 2013.

J Paisley, C Wang, D Blei. "The discrete infinite logistic normal distribution," *Bayesian Analysis*, v.7, 2012, p. 235.

M Hoffman, D Blei, J Paisley, C Wang. "Stochastic Variational Inference," *Journal of Machine Learning Research*, v.NA, 2013.

R. Waeber, PI Frazier, SG Henderson. "Bisection Search with Noisy Responses," *SIAM Journal on Control and Optimization*, v.51, 2013, p. 2261.

J. Paisley, C. Wang, D. Blei, and M. Jordan.. "A nested HDP for hierarchical topic modeling.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press., 2015. 등 14편

[NSF 빅데이터 예비후보03]

User Driven Architecture for Data Discovery

CIF21 DIBBs:

Award Number:1443070; Principal Investigator:Giridhar Manepalli; Co-Principal Investigator:Laurence Lannom, Allison Powell; Organization:Corporation for National Research Initiatives (NRI);NSF Organization:OAC Start Date:09/01/2014; Award Amount:\$1,484,940.00; ~April 30, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1443070&HistoricalAwards=false

The number, size, and availability of scientific datasets have grown enormously over the last few years. As scientific activity becomes more data intensive and collaborative, a key challenge for cross-disciplinary research will be discovery of diverse data sets, managed within distributed repositories and registries. Currently, discovery of information on the Internet is largely performed through automated approaches, characterized by web crawling and associated algorithms, or labor intensive indexing and categorization, such as the National Library of Medicine index for medical literature. There are significant amounts of data housed in repositories where only researchers with expertise in the specific field know and access the data.

This project builds a **user driven architecture for data discovery (UDADD)**, a capability that enhances discovery of scientific datasets by **building a global index from diverse communities** with minimal input. In the UDADD approach user actions, such as dataset queries or downloads, drive the construction of a global index. These actions are recorded and gathered automatically, through cooperation with repository managers. Two software plugins are provided to help the repositories interact with the UDADD system. The architecture includes ranking techniques based on frequency and recency of use of the datasets.

The pilot architecture will be demonstrated and evaluated using cooperating repositories within the DataNet Federation Consortium. Currently, six science and engineering communities participate in the consortium, including national scale projects in oceanography, social science, cognitive science, hydrology, engineering, and plant biology.

[NSF 빅데이터 예비후보04]

Human and Machine Co-Processing

BIGDATA: F: DKA: CSD:

Award Number:1447449; Principal Investigator:Robert Nowak; Co-Principal Investigator:Stephen Wright, Rebecca Willett; Organization:University of Wisconsin-Madison;NSF Organization:IIS Start Date:09/01/2014; Award Amount:\$1,396,830.00; ~August 31, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1447449&HistoricalAwards=false

Human experts are crucial to data analysis. Their roles include sifting through large datasets to facilitate search, retrieval, and machine learning. Humans often perform much better than machines at such tasks, but the speed and capacity of human experts is a limiting factor in the human-machine co-processing. This project is addressing two aspects of human-machine co-processing: **winnowing Big Data** to produce manageable subsets for human expert analysis, and **machine learning algorithms that learn efficiently from human experts with a minimal amount of human interaction**. This has a wide range of applications; to ensure broad applicability of the results the project is evaluating the techniques in multiple domains: cognitive science, large-scale astronomical data analysis, and experimental design in materials science.

The approach used for data winnowing is based on developing predictive models and identifying data that does not fit the models. A key research challenge is non-stationary environments: the underlying model changes over time. Preliminary work shows promise on selection from a finite set of models, and new work investigates more flexible parametric models. The active learning task uses the multi-armed bandit problem to model identify which features have the greatest impact on human decisions. This task also investigates learning from comparisons/rankings rather than predictions; conjecturing that there may exist low-dimensional structure governing human reasoning and decision-making that enables learning with significantly fewer comparisons than might otherwise be required. A common theme in both tasks is ensuring computational complexity is low enough to facilitate real-time interactions with human experts in spite of the volume of data. This is achieved using bounded approximations and convex relaxations of the optimization programs used to guide the interaction.

E. Hall and R. Willett. "Online Convex Optimization in Dynamic Environments," IEEE Journal of Selected Topics in Signal Processing, Special Issue on Big Data, 2015. doi:10.1109/JSTSP.2015.2404790

Stephen J. Wright. "Coordinate Descent Algorithms," Mathematical Programming, Series B, v.151, 2015, p. 3.

Eric Hall, Garvesh Raskutti, and Rebecca Willett. "Inferring High-Dimensional Poisson Autoregressive Point Processes," IEEE Workshop on Statistical Signal Processing, 2016.

Jun, K.S., Jamieson, K., Nowak, R. and Zhu, X.. "Top Arm Identification in Multi-Armed Bandits with Batch Arm Pulls.," In The 19th International Conference on Artificial Intelligence and Statistics (AISTATS)., 2016.

Rau, M. A., Mason, B., & Nowak, R.. "How to model implicit knowledge? Use of metric learning to assess student perceptions of visual representations.," In T. Barnes, M. Chi & M. Feng (Eds.), Proceedings of the 9th International Conference on Educational Data Mining (pp. 199-206). Raleigh NC., 2016.

Davis Gilton and Rebecca Willett. "Sparse Linear Contextual Bandits via Relevance Vector Machines," 12th International Conference on Sampling Theory and Applications, 2017.

Kennedy, R., Balzano, L., Wright, S. J., and Taylor, C. J.. "Online algorithms for factorization-based structure from motion," Computer Vision and Image Understanding, v.150, 2016, p. 139.

Kwang-Sun Jun, Francesco Orabona, Rebecca M. Willett, and Stephen J. Wright. "Improved strongly adaptive online learning using coin betting," Journal, 2017.

Nikhil Rao, Ravi Ganti, Laura Balzano, Rebecca Willett, and Robert Nowak. "On learning high-dimensional structured single index models," AAAI Conference on Artificial Intelligence, 2017.

Rebecca Willett. "Signal Representations in Modern Signal Processing," IEEE International Conference on Acoustics, Speech, and Signal Processing, 2017. 등 11편

[NSF 빅데이터 예비후보05]

Big Data Modeling and Analysis with Depth and Scale

BIGDATA: F: DKM: DKA:

Award Number:1447549; Principal Investigator:Coimbatore Ramakrishnan; Co-Principal Investigator:Yanhong Liu, Scott Smolka, I. Ramakrishnan, Maureen O'Leary; Organization:SUNY at Stony Brook;NSF Organization:IIS Start Date:08/01/2014; Award Amount:\$1,500,000.00; ~July 31, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1447549&HistoricalAwards=false

An important step in understanding large volumes of data is the construction of a model: a succinct but abstract representation of the phenomenon that produced the data. In order to understand a phenomenon, a data analyst needs to be able to propose a model, evaluate how the proposed model explains the data, and refine the model as new data becomes available. Statistical models, which specify relationships among random variables, have traditionally been used to understand large volumes of noisy data. Logical models have been used widely in databases and knowledge bases for organizing and reasoning with large and complex data sets. This project is aimed at developing **a programming language and system for the creation, evaluation and refinement of combined statistical and logical models for the express purpose of understanding very large and complex data sets**. Apart from its direct effect on model development for Big Data problems, the semantic foundations and scalable computing infrastructure resulting from this project is expected to directly impact the areas of system development and verification, planning, and optimization, with broad application in Science and Engineering. The tools developed in this project will facilitate the training of a new generation of scientists capable of transforming data into knowledge for use across disciplines. The project's education and outreach component is designed to train select undergraduate students on Big Data modeling and analysis via annual workshops and research mentorship; and graduate students via curriculum modifications including a specialization in Data Science.

The project will develop Px, a language with well-defined declarative semantics, to support high-level model construction and analysis. Px will be capable of expressing generative and discriminative probabilistic and relational models, and the Px system will support complex queries over such models. The project will encompass three significant and complementary research directions, aimed at developing: (1) semantic foundations, including language constructs needed for succinct specification of complex models with rich logical and statistical structure; (2) scalable inference techniques combining exact and approximate methods, and query optimizations over combined logic/statistical models; and (3) programming extensions as well as static and dynamic analyses to support the creation and refinement of complex models. The Px language and system will be evaluated using two important and diverse application problems: (1) analysis and verification of infinite-state probabilistic systems, including parameterized systems, and (2) construction of phylogenetic trees from phenomic data, used in the Tree of Life project, for mapping the evolutionary history of organisms. The project is expected to make significant contributions towards creating a unifying framework combining probabilistic inference, logical inference, and constraint processing, with an emphasis on semantic clarity, efficiency, and scalability. The project will also demonstrate the practical utility of the proposed integrated framework by developing complex models from big data that take advantage of this technology in fundamental ways.

Arun Nampally, C. R. Ramakrishnan. "Constraint-Based Inference in Probabilistic Logic Programs.," *International Workshop on Probabilistic Logic Programming*. Cork, Ireland., 2015, p. 46.

Jon Brandvein and Yanhong A. Liu. "Removing Runtime Overhead for Optimized Object Queries," *The ACM SIGPLAN Workshop on Partial Evaluation and Program Manipulation*, 2016. 등 21편

[NSF 빅데이터 예비후보06]

Addressing the two V's of Veracity and Variety in Big Data

BIGDATA: F: DKM:

Award Number:1447795; Principal Investigator:Nitesh Chawla; Co-Principal Investigator:Dong Wang, Thanuka Wickramaratne; Organization:University of Notre Dame;NSF Organization:IIS Start Date:09/01/2014; Award Amount:\$1,000,000.00; Relevance:72.0; ~August 31, 2018 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1447795&HistoricalAwards=false

Data of questionable quality have led to significantly negative economic and social impacts on organizations, leading to overrun in costs, lost revenue, and decreased efficiencies. The issues on data reliability, credibility, and provenance have become even more daunting when dealing with the variety of data, especially data that are not directly collected by an organization, but from the third-party sources such as social media, data brokers, and crowdsourcing. To address such issues, this project aims to develop a **Data Valuation Engine (DVE)** that solves the critical problem of data reliability, credibility and provenance, and provides accountability and quality processes right from data acquisition. The DVE leverages and innovates techniques in estimation theory, data fusion and machine learning to fill a critical gap in data accountability and quality, thereby providing a transformative step in countering the ubiquitous data quality issues found in almost every application domain from business to environment to health to national security. The DVE will be integrated in the Hadoop ecosystem and will be agnostic to the data source, application or analytics, and provided as a hosted solution to the community. The user will interact with DVE by providing the data sources and relevant data necessary to solve a problem.

The DVE in this project will be developed in a largely application-independent manner. The key challenges to develop this engine include: (i) How to generate the data quality indication labels to score data sources and the content of data based on various factors such as reliability, credibility, uncertainty and confidence? (ii) How to integrate data from various sources with different labeled scores? (iii) How to robustly evaluate the proposed engine in a broad spectrum of applications that serve as a proxy of a variety of real-world scenarios? The research plan has been designed to synergistically address the above challenges with a robust evaluation plan. Given the generality of the proposed methods, models and system, the project will potentially impact variety of applications of science, engineering, and social science and have broad environmental, economic, and health benefits. The PIs will release open source software and applicable data. The PIs will also provide a hosted DVE platform for a broad user and participant base. This project is also providing students with greater exposure to the areas of big data analytics, cloud computing, data fusion and data mining, both in courses and research experience

Keith Feldman, Nicholas Hazecamp, and Nitesh V. Chawla. "Mining the Clinical Narrative: All Text Are Not Equal," IEEE International Conference on Health Informatics, 2016.

Chao Huang and Dong Wang. "On Interesting Place Finding in Social Sensing: An Emerging Smart City Application Paradigm," IEEE International Conference on Smart City, 2015.

Chao Huang and Dong Wang. "Unsupervised Interesting Places Discovery in Location-Based Social Sensing," IEEE International Conference on Distributed Computing in Sensor Systems, 2016. 등 12편

[NSF 빅데이터 예비후보07]

Large-Scale Transductive Learning from Heterogeneous Data Sources

BIGDATA: F:

Award Number:1546329; Principal Investigator:Yiming Yang; Co-Principal Investigator;; Organization:Carnegie-Mellon University;NSF Organization:IIS Start Date:01/01/2016; Award Amount:\$1,204,545.00; Relevance:72.0; ~December 31, 2019 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1546329&HistoricalAwards=false

Important problems in the big-data era involve predictions based on **heterogeneous sources of information and the dependency structures in data**. In recommendation systems, for example, predictions need to be made not only based on observed user ratings over items (movies, books, music, shopping products, etc.), but also based on information such as demographical data of users and textual descriptions of items. In event detection from textual data (news stories, tweets, maintenance reports, legal documents, etc.), joint inference must be based on who (agents), what (event types or topics), where (locations) and when (dates), and also based on the connections among agents (in social networks), topics (in an event-type ontology), locations (in a map) and temporal co-occurrences. The fundamental research questions therefore include: (1) how to develop a unified optimization framework for predictions based on heterogeneous information and dependency structures in various kinds of tasks; (2) how to make the inference computationally tractable when the combined space of model parameters is extremely large; and (3) how to significantly enhance the prediction power of the system by leveraging massively available unlabeled data in addition to human-annotated training data which are often sparse.

This project will address the three challenges via the following four approaches.

(1) A unified representation of heterogeneous information sources using **product graphs**: This framework aims to represent heterogeneous sources of data and intra-source dependencies, such as social connections among users, semantic similarities among items, contextual correlations among keywords, topical similarities among documents, hierarchical relations among topic labels, and so on. Each data source will be represented using a graph, and the individual graphs of multiple sources will be combined into a product graph where each node corresponds to a tuple of nodes in the individual graphs, and each link aggregates the links in the individual graphs.

(2) Transductive learning over graph products: This project plans to reduce the inference problems in a broad range of prediction tasks to semi-supervised transductive learning problems over the product graphs mentioned above. The training data in each task (of classification, regression or link prediction) will be represented as a subset of labeled (or scored) nodes in the product graph, and the labels (or scores) of those nodes will be propagated over the links in the product graph until convergence. This project will study various kinds of graph transductions theoretically and empirically.

(3) Large-scale optimization algorithms: The induced product graphs are typically extremely large. To address the computational bottlenecks, this project will develop new scalable algorithms based on theoretical properties and computational characteristics of spectral graph products, including adapted versions of rank-reduced matrix factorization, aggressive basis pruning, and sampling-based low-rank approximation.

(4) Thorough evaluations in multiple important applications: The proposed new approach will be evaluated on benchmark data collections for context-aware collaborative filtering, semi-structured event detection and tracking, and expert finding via multi-source social network analysis.

The proposed work, if successful, will offer principled solutions for enhancing the prediction power of systems in a broad range of tasks, whenever recommendation, classification and regression are involved. Technical impacts of the proposed work are expected in multiple research fields. For further information see the project web site at: <http://nyc.lti.cs.cmu.edu/gp-trans/index.html>

Hanxiao Liu and Yiming Yang. "Cross-Graph Learning of Multi-Relational Associations," International Conference on Machine Learning (ICML), 2016.

Xu, Ruochen and Yang, Yiming. "Cross-lingual Distillation for Text Classification," Annual Meeting of the Association for Computational Linguistics, 2017. 11편

[NSF 빅데이터 예비후보08]

Study of a Cyber-Enabled Social Computing Framework for Improving Practice in Online Computing Communities

BIGDATA: Collaborative Research: F:

Award Number:1546393; Principal Investigator:Carolyn Rose; Co-Principal Investigator:James Herbsleb; Organization:Carnegie-Mellon University;NSF Organization:IIS Start Date:01/01/2016; Award Amount:\$1,598,501.00; Relevance:72.0; December 31, 2019 (Estimated)

Data science techniques have revolutionized many academic fields and led to terrific gains in the commercial sector. They have to date been underutilized in solving critical problems in the US educational system, particularly in understanding Science, Technology, Engineering and Mathematics (STEM) learning and learning environments, broadening participation in STEM, and increasing retention for students traditionally underserved in STEM. The goals of the Directorate for Education and Human Resources through the Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering (BIGDATA) program are to advance fundamental understanding of key questions in the field, and catalyze the use of data science in Education Research. Computing has become an integral part of the practice of in modern science, technology engineering, and mathematics (STEM) fields. As a result, the *STEM+Computing Partnership (STEM+C) program* seeks to integrate the use of computational approaches in STEM teaching and learning and understand how this integration can improve STEM learning, engagement, and persistence. In this proposal, the Principal Investigators (PIs) will examine environments that many people engage in independently to learn computing, online communities and massive open online courses (MOOCs). This activity could be very compelling as people come to these environments because they have personal goals to learn the material. However, a challenge in these environments is that there is little support for learning. In addition, these environments are not adaptive to learners' needs. This project will tackle both of these challenges. The PIs will first characterize groups of learners to understand their needs and then design approaches to personalizing the environments based on those needs.

The PIs address a need for both learning and independent online work communities by providing a combined learning work community. Many authentic production communities, such as **GitHub**, do not provide support for novices who want to learn to contribute. Similarly many learning communities, such as **MOOCs** and other online learning environments, do not provide outlets for learners' products to become authentic. The PIs will **combine data from MOOCs and online communities to discover groups of participants who behave in similar ways and investigate how to support the needs of these groups**. This will lead to the proposed novel combined learning and work community that both provides support and offers authentic outlets for work products that are valued beyond a particular course.

Trainer, Erik H. and Kalyanasundaram, Arun and Herbsleb, James D.. "E-Mentoring for Software Engineering: A Socio-Technical Perspective," International Conference on Software Engineering: Software Engineering and Education Track, 2017. doi:10.1109/ICSE-SEET.2017.19 Citation details

Filippova, Anna and Trainer, Erik and Herbsleb, James D.. "From Diversity by Numbers to Diversity as Process: Supporting Inclusiveness in Software Development Teams with Brainstorming," International Conference on Software Engineering, 2017. doi:10.1109/ICSE.2017.22 Citation details

Herbsleb, James. "Building a socio-technical theory of coordination: why and how (outstanding research award)," ACM SIGSOFT International Symposium on Foundations of Software Engineering, 2016. doi:10.1145/2950290.2994160 Citation details

Bogart, Christopher and Kästner, Christian and Herbsleb, James and Thung, Ferdian. "How to break an API: cost negotiation and community values in three software ecosystems," ACM SIGSOFT International Symposium on Foundations of Software Engineering, 2016. doi:10.1145/2950290.2950325 Citation details

[NSF 빅데이터 예비후보09]

Hype Cycles of Scientific Innovation

BIGDATA: IA:

Award Number:1633036; Principal Investigator:Daniel McFarland; Co-Principal Investigator:Daniel Jurafsky; Organization:Stanford University;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$1,253,468.00; Relevance:72.0; ~August 31, 2019 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1633036&HistoricalAwards=false

1990-2016 사이에 발표된 도서, 과학 논문, 정부 지원 프로젝트, 특허 등의 텍스트 데이터를 수집하여 주제 분석 (topic analysis) 등 실시

Scientific discovery is a collective process based on collaboration, assessment and consensus. But the enormous expansion of scientific research makes it difficult to tell which intellectual efforts forge collective advances. Better models for identifying and tracking scientific movements from the vast collections of articles, books, grants, and patents that comprise scientific and academic work is crucial to improving our nation's ability to make advances in science and industry, helping **policy makers, funding agencies, and venture capitalists** as well as **scientists and scholars themselves**. Our project studies **innovation by collecting and analyzing texts in vast collections of scientific and other scholarly articles, books, grants and patents**. By looking at the subtle patterns of language and how they change over time, we can describe and predict where and when collective trends of knowledge innovation emerge and decline; which scientific ideas and movements result in translational knowledge key to industry and health care; and what the key drivers are for such collective intellectual movements. This work is helping identify where the potential is greatest for innovation, which fields are most primed and receptive to the arrival of new discoveries, and the times and places where resources have the greatest influence.

The project is based on a large dataset of texts the researchers have compiled on **US research activity from 1990-2016**, including **scientific articles, grants, and patents**, and by disambiguating and linking mentions of individual people across these datasets. This project uses **topic modeling** and other **natural language processing** algorithms to identify distinct intellectual movements in these corpora by drawing on the sub-languages that characterize them, also applying methods to validate and evaluate these movements. This allows the researchers to model the trajectories of movements over time with tools like latent growth mixture modeling and k-spectral clustering, identifying the mechanisms associated with each movement, how they rise and fall over time, and how they translate into industrial or health applications. For example, the project studies how the trajectory of intellectual movements is a function of their environment and competing research efforts, and show the ways in which it depends on the timing and magnitude of key resources (e.g., money, recognition, new recruits and trainees, social networks of support, or the coherence of the knowledge itself). In such a fashion, this work is an important first step in unraveling the recipe for innovation as a collective, episodic process.

[NSF 빅데이터 예비후보10]

OSCAR - Open Source Supply Chains and Avoidance of Risk: An Evidence Based Approach to Improve FLOSS Supply Chains

BIGDATA: Collaborative Research: IA:

Award Number:1633437; Principal Investigator:Audris Mockus; Co-Principal Investigator:Randy Bradley, Bogdan Bichescu, Russell Zaretski; Organization:University of Tennessee Knoxville;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$1,300,000.00; Relevance:72.0; ~August 31, 2020 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1633437&HistoricalAwards=false

Open source software is an engine for innovation and a critical infrastructure for the nation and yet it is implemented by communities formed from a loose collection of individuals. With each software project relying on thousands of other software projects, this complex and dynamic supply chain introduces new risks and unpredictability, since, unlike in traditional software projects, no contractual relationships with the community exist and individuals could simply lose interest or move on to other activities.

The big data-based approach to software supply chains will stimulate academic and practical work. The tools and practices to quantify and mitigate risks in the rapidly changing global environment with no centralized control or authority will lead to dramatic reductions in risk manifested in, for example, the spread of vulnerabilities thus making the nation both safer and more innovative. The theoretical frameworks and approaches developed will likely influence research and practice in other supply chain contexts.

The objective of this research is to advance the state of knowledge of software supply chains by **collecting and integrating massive public operational data representing development activity and source code from all open source projects** and **using it to develop novel theories, methods, and tools**. The construction and analysis of the entire open source supply chain provides static and dynamic properties of the network, risk propagation, and system-level risks. Novel statistical and game-theoretic models are used to assess and mitigate these risks, while methods to contextualize, augment, and correct operational data provide ways to cope with data's size, complexity, and observational nature.

<부록 14> NSF HPC 예비 후보

[NSF HPC 예비 후보01]

Compute on Data Path: Combating Data Movement in High Performance Computing

SHF: Medium:

Award Number:1409946; Principal Investigator:Yong Chen; Co-Principal Investigator:Robert Ross, Dries Kimpe, Yonghong Yan, Barbara Chapman; Organization:Texas Tech University;NSF Organization:CCF
Start Date:06/01/2014; Award Amount:\$1,016,000.00; ~May 31, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1409946&HistoricalAwards=false

High performance computing enabled simulation has been widely considered a third pillar of science along with theory and experimentation, and is a strategic tool in many aspects of scientific discovery and innovation. High performance computing simulations, however, have become highly data intensive in recent years due to data acquisition and generation becoming much cheaper, newer high-resolution multi-model scientific discovery producing and requiring more data, and the insight that useful data can be mined out of large amounts of data being substantially increased.

This project combats the increasingly critical data movement challenge in high performance computing. This project studies **the feasibility of a new Compute on Data Path methodology** that expects to improve the performance and energy efficiency for high performance computing. This new methodology models both computations and data as objects with a data model that encapsulates and binds them. It fuses data motion and computation leveraging programming model and compiler. It develops an object-based store and runtime to enable computations along data path pipeline. In recent years, a proliferation of advanced high performance computing architectures including multi- and many-core systems, co-processors and accelerators, and heterogeneous computing platforms have been observed. The software solution that addresses the critical data movement challenge, however, has significantly lagged behind. This project has the potential of advancing the understandings and the software solution and further unleashing the power of high performance computing enabled simulation.

D. Khaldi, S. Ge, D. Eachempati, P. Jouvelot and B. Chapman. "A Team-Based Methodology of Memory Hierarchy-Aware Runtime Support in Coarray Fortran," *IEEE Cluster 2015*, 2015, p. 448. [doi:2015 International Conference on Cluster Computing](#)

Dounia Khaldi, Pierre Jouvelot, Francois Irigoien, Corinne Ancourt and Barbara Chapman. "LLVM Parallel Intermediate Representation: Design and Evaluation using OpenSHMEM Communications," *Second Workshop on the LLVM Compiler Infrastructure in HPC (LLVM-HPC2, held in conjunction with SC15)*, 2015. [doi:978-1-4503-4005-2](#)

Dong Dai, Yong Chen, Philip Carns, Jonathan Jenkins, Wei Zhang and Robert Ross. "GraphMeta: A Graph-based Engine for Managing Large-Scale HPC Rich Metadata," *The 45th International Conference on Parallel Processing (ICPP-2016)*, 2016.

Dong Dai; Philip Carns; Robert Ross; John Jenkins; Nicholas Muirhead; Yong Chen. "An Asynchronous Traversal Engine for Graph-based Rich Metadata Management," *International Journal on Parallel Computing*, 2016.

Dounia Khaldi and Barbara Chapman. "Towards an Automatic HBM Allocation using LLVM: Case Study with Knights Landing," *Proceedings of the Third Workshop on LLVM Compiler Infrastructure in HPC*, 2016, p. 12. [doi:978-1-5090-3878-7](#)

Robert Latham, Matthieu Dorier, and Robert Ross. "Get Out of the Way! Applying Compression to Internal Data Structures," *Proceedings of the PDSW-DISCS 2016 workshop*, 2016.

Solmaz Salehian, Jiawen Liu and Yonghong Yan. "Comparison of Threading Programming Models," *International Workshop on High-level Parallel Programming Models and Supportive Environments, in conjunction with IPDPS 2017*, 2017.

Yonghong Yan, Jiawen Liu, Kirk W. Cameron and Mariam Umar. "HOMP: Automated Distribution of Parallel Loops and Data in Highly Parallel Accelerator-Based Systems," *IPDPS 2017*, 2017.

[NSF HPC 예비후보02]

Middleware and High Performance Analytics Libraries for Scalable Data Science

CIF21 DIBBs:

Award Number:1443054; Principal Investigator:Geoffrey Fox; Co-Principal Investigator:Judy Qiu, Shantenu Jha, Fusheng Wang, Madhav Marathe; Organization:Indiana University;NSF Organization:OAC Start Date:10/01/2014; Award Amount:\$5,154,258.00; September 30, 2019 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1443054&HistoricalAwards=false

Many scientific problems depend on the ability to analyze and compute on large amounts of data. This analysis often does not scale well; its effectiveness is hampered by the increasing volume, variety and rate of change (velocity) of big data. This project will design, develop and implement building blocks that enable a fundamental improvement in the ability to support data intensive analysis on a broad range of cyberinfrastructure, including that supported by NSF for the scientific community. The project will integrate features of traditional high-performance computing, such as scientific libraries, communication and resource management middleware, with the rich set of capabilities found in the commercial Big Data ecosystem. The latter includes many important software systems such as Hadoop, available from the Apache open source community. A collaboration between university teams at Arizona, Emory, Indiana (lead), Kansas, Rutgers, Virginia Tech, and Utah provides the broad expertise needed to design and successfully execute the project. The project will engage scientists and educators with annual workshops and activities at discipline-specific meetings, both to gather requirements for and feedback on its software. It will include under-represented communities with summer experiences, and will develop curriculum modules that include demonstrations built as 'Data Analytics as a Service.'

The project will design and implement a **software Middleware for Data-Intensive Analytics and Science (MIDAS)** that will enable scalable applications with the performance of HPC (High Performance Computing) and the rich functionality of the commodity Apache Big Data Stack. Further, this project will design and implement a set of cross-cutting high-performance data-analysis libraries; **SPIDAL (Scalable Parallel Interoperable Data Analytics Library)** will support new programming and execution models for data-intensive analysis in a wide range of science and engineering applications. The project addresses major data challenges in seven different communities: Biomolecular Simulations, Network and Computational Social Science, Epidemiology, Computer Vision, Spatial Geographical Information Systems, Remote Sensing for Polar Science, and Pathology Informatics. The project libraries will have the same beneficial impact on data analytics that scientific libraries such as PETSc, MPI and ScaLAPACK have had for supercomputer simulations. These libraries will be implemented to be scalable and interoperable across a range of computing systems including clouds, clusters and supercomputers.

F. Wang, A. Aji and H. Vo. "High Performance Spatial Queries for Spatial Big Data: from Medical Imaging to GIS.," *ACM Special Interest Group on Spatial Information-- Special Issue*, v.6, 2014. [doi:10.1145/2766196.2766199](https://doi.org/10.1145/2766196.2766199)

A. Luckow, P. Mantha, S. Jha. "Pilot-Abstraction: A Valid Abstraction for Data-Intensive Applications on HPC, Hadoop and Cloud Infrastructures?," *IPDPS*, 2015.

Bhuiyan, Hasanuzzaman, Maleq Khan, Jiangzhuo Chen, and Madhav Marathe. "Parallel algorithms for switching edges in heterogeneous graphs," *Journal of Parallel and Distributed Computing*, v.104, 2017, p. 19-35.

X. Sun, W. Pittard, T. Xu, Tianlei, M. Zwick, X. Jiang, F. Wang, and Z. Qin. "OmicSeq: A web-based search engine for exploring omics datasets," *Nucleic Acids Research*, 2017. [doi:10.1093/nar/gkx258](https://doi.org/10.1093/nar/gkx258)

Z Heidari, DR Roe, R Galindo-Murillo, JB Ghasemi, and TE Cheatham III. "Using Wavelet Analysis To Assist in Identification of Significant Events in Molecular Dynamics Simulations," *Journal of Chemical Information and Modeling*, v.56, 2016, p. 1282-1291.

Geoffrey C. Fox, Shantenu Jha, Judy Qiu, and Andre Luckow. "Ogres: A Systematic Approach to Big Data Benchmarks," *Big Data and Extreme-scale Computing (BDEC)*, 2015.

[NSF HPC 예비후보03]

Performance Application Programming Interface for Extreme-Scale Environments (PAPI-EX)

SI2-SSI: Collaborative Proposal:

Award Number:1450429; Principal Investigator:Jack Dongarra; Co-Principal Investigator:Heike Jagode, Anthony Danalis; Organization:University of Tennessee Knoxville;NSF Organization:OAC Start Date:09/01/2015; Award Amount:\$2,126,446.00; Relevance:67.16; August 31, 2019 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1450429&HistoricalAwards=false

Modern High Performance Computing (HPC) systems continue to increase in size and complexity. Tools to measure application performance in these increasingly complex environments must also increase the richness of their measurements to provide insights into the increasingly intricate ways in which software and hardware interact. **The PAPI performance-monitoring library** has provided a clear, portable interface to the hardware performance counters available on all modern CPUs and some other components of interest (scattered across the chip and system). Widely deployed and widely used, PAPI has established itself as fundamental software infrastructure, not only for the scientific computing community, but for many industry users of HPC as well. But the radical changes in processor and system design that have occurred over the past several years pose new challenges to PAPI and the HPC software infrastructure as a whole. The PAPI-EX project **integrates critical PAPI enhancements** that flow from both governmental and industry research investments, focusing on processor and system design changes that are expected to be present in every extreme scale platform on the path to **exascale computing**.

The primary impact of PAPI-EX is a direct function of the importance of the PAPI library. PAPI has been in predominant use by tool developers, major national HPC centers, system vendors, and application developers for over 15 years. PAPI-EX builds on that foundation. As important research infrastructure, the PAPI-EX project allows PAPI to continue to play its essential role in the face of the revolutionary changes in the design and scale of new systems. In terms of enhancing discovery and education, the list of partners working with PAPI-EX includes NSF computing centers, major tool developers, major system vendors, and individual community leaders, and this diverse group will help facilitate training sessions, targeted workshops, and mini-symposia at national and international meetings. Finally, the active promotion of PAPI by many major system vendors means that PAPI, and therefore PAPI-EX, will continue to deliver major benefits for government and industry in many domains.

PAPI-EX addresses a hardware environment in which the cores of current and future multicore CPUs share various performance-critical resources (a.k.a., **'inter-core' resources**), including power management, on-chip networks, the memory hierarchy, and memory controllers between cores. Failure to manage contention for these 'inter-core' resources has already become a major drag on overall application performance. Consequently, the lack of ability to reveal the actual behavior of these resources at a low level, has become very problematic for the users of the many performance tools (e.g., TAU, HPCToolkit, Open|SpeedShop, Vampir, Scalasca, CrayPat, Active Harmony, etc.). PAPI-EX enhances and extends PAPI to solve this critical problem and prepare it to play its well-established role in HPC performance optimization. Accordingly, PAPI-EX targets the following objectives: (1) Develop shared hardware counter support that includes system-wide and inter-core measurements; (2) Provide support for data-flow based runtime systems; (3) Create a sampling interface to record streams of performance data with relevant context; (4) Combine an easy-to-use tool for text-based application performance analysis with updates to PAPI's high-level API to create a basic, 'out of the box' instrumentation API.

[NSF HPC 예비후보04] Expanding Volunteer Computing

Collaborative Research: SI2-SSI:

Award Number:1664022; Principal Investigator:Lucas Wilson; Co-Principal Investigator;; Organization:University of Texas at Austin;NSF Organization:OAC Start Date:05/15/2017; Award Amount:\$500,000.00; April 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1664022&HistoricalAwards=false

Volunteer computing (VC) uses **donated computing time consumer devices** such as home computers and smartphones to do scientific computing. It has been shown that VC can provide greater computing power, at lower cost, than conventional approaches such as organizational computing centers and commercial clouds. **BOINC** is the most common software framework for VC. Essentially, donors of computing time simply have to load BOINC on their computer or smartphone, and then register to donate at the BOINC web site. VC provides "high throughput computing": handling lots of independent jobs, with performance goals based on the rate of job completion rather than completion time for individual jobs. This type of computing (all known as high-throughput computing) is in great demand in most areas of science. Until now, the adoption of VC has been limited by its structure. For example, VC projects (such as **Einstein@home** and **Rosetta@home**) are operated by individual research groups, and volunteers must browse and choose from among many such projects. As a result, there are relatively few VC projects, and volunteers are mostly tech-savvy computer enthusiasts. This project aims to solve these problems using two complementary development efforts: First, it will add BOINC-based VC conduits to two major high-performance computing providers: (a) the Texas Advanced Computing Center, a supercomputer center, and (b) nanoHUB, a web portal for nano science that provides computing capabilities. Also, a unified control interface to VC will be developed, tentatively called Science United, where donors can register. The project will benefit thousands of scientists who use these facilities, and it will create technology that makes it easy for other HPC providers to add their own VC back ends. Also, Science United will provide a simpler interface to BOINC volunteers where they will register to support scientific areas, rather than specific projects. Science United will also serve as an allocator of computing power among projects. Thus, new projects will no longer have to do their own marketing and publicity to recruit volunteers. Finally, the creation of a single VC "brand" (i.e Science United) will allow coherent marketing of VC to the public. By creating a huge pool of low-cost computing power that will benefit thousands of scientists, and increasing public awareness of and interest in science, the project plans to establish VC as a central and long-term part of the U.S. scientific cyber infrastructure.

Adding VC to an existing HPC facility involves several technical issues, which will be addressed as follows: (1) Packaging science applications (which typically run on Linux cluster nodes) to run on home computers (mostly Windows, some Mac and Linux): the team is developing an approach using VirtualBox and Docker, in which the application and its environment (Linux distribution, libraries, executables) are represented as a set of layers comprising a Docker image, which is then run as a container within a Linux virtual machine on the volunteer device. This has numerous advantages: it reduces the work of packaging applications to near zero; it minimizes network traffic because a given Docker layer is downloaded to a host only once; and it provides a strong security sandbox so that volunteer computers are protected from buggy or malicious applications, (2) File management: Input and output files must be moved between existing private servers and public-facing servers that are accessible to the outside Internet. A file management system will be developed, based on Web RPCs, for this purpose. This system will use content-based naming so that a given file is transferred and stored only once. It also maintains job/file associations so that files can be automatically deleted from the public server when they are no longer needed. (3) Submitting and monitoring jobs: BOINC provides a web interface for efficiently submitting and monitoring large batches of jobs. These were originally developed as part of a system to migrate HTCCondor jobs to BOINC. This project is extending it to support the additional requirements of TACC and nanoHUB. Note that these new capabilities are not specific to TACC or nanoHUB: they provide the glue needed to easily add BOINC-based VC to any existing HTC facility. The team is also developing RPC bindings in several languages (Python, C++, PHP). The other component of the project, Science United, is a database-driven web site and an associated web service for the BOINC clients. Science United will control volunteer hosts (i.e. tell them which projects to work for) using BOINC's "Account Manager" mechanism, in which the BOINC client on each host periodically contacts Science United and is told what projects to run. Project servers, not Science United, will distribute jobs and files. Science United will define a set of "keywords" for science areas (physics, biomedicine, environment, etc.) and for location (country, institution). Projects will be labelled with appropriate keywords. Volunteers will have a yes/no/maybe interface for specifying the types of jobs they want to run. Science United will thus provide a mechanism in which a fraction of total computing capacity can be allocated to a project for a given period. Because total capacity changes slowly over time, this allows near-certain guaranteed allocations. Science United will embody a scheduling system that attempts to enforce allocations, honor volunteer preferences, and maximize throughput. Finally, Science United will do detailed accounting of computing. Volunteer hosts will tell Science United how much work (measured by CPU time and FLOPs, GPU time and FLOPs, and number of jobs) they have done for each project. Science United will maintain historical records of this data for volunteers and projects, and current totals with finer granularity (e.g. for each host/project combination). Finally, Science United will provide web interfaces letting volunteers see their contribution status and history, and letting administrators add projects, control allocations, and view accounting data.

[NSF HPC 예비후보05]

Building High-Availability Data Capabilities in Data-Centric Cyberinfrastructure

CIF21 DIBBs: PD:

Award Number:1724845; Principal Investigator:Haiying (Helen) Shen; Co-Principal Investigator;; Organization:University of Virginia Main Campus;NSF Organization:OAC Start Date:08/01/2017; Award Amount:\$500,000.00; Relevance:65.61; July 31, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1724845&HistoricalAwards=false

This project supports data-related analysis in a wide range of science and engineering applications. It will contribute to the development of scalable data-centric cyberinfrastructure capabilities, to accelerate interdisciplinary and collaborative research. The **exascale file system** will serve as catalyst for research in data storage architectures, and will enable new data-focused services and capabilities that advance scientific discoveries, collaborations, and innovations. The project addresses a major data challenge common to a range of communities such as social science, economics, and bioengineering, and will provide **thorough training** and **collaborative research** opportunities for project participants.

Both high performance computing (HPC) clusters and Hadoop clusters use file systems. A Hadoop cluster uses the Hadoop Distributed File System (HDFS) that resides on compute nodes, while an HPC cluster usually uses a remote storage system. Despite years of efforts on research and application development on HPC and Hadoop clusters, the file systems in both types of clusters still face a formidable challenge, that of achieving exascale computing capabilities. The centralized data indexing in HDFS and HPC storage architectures cannot provide **high scalability** and **reliability**, and both HDFS and HPC storage architectures have shortcomings such as **single point of failure** and **insufficiently efficient data access**. This project builds **scalable high-availability data capabilities in data-centric cyberinfrastructure** to overcome the shortcomings and create a highly scalable file system with new techniques for distributed load balancing, data replication and consistency maintenance.

[NSF HPC 예비후보06]

Hierarchical Tuning of Floating-Point Computations

SHF: Medium:

Award Number:1704715; Principal Investigator:Ganesh Gopalakrishnan; Co-Principal Investigator:Hari Sundar, Mary Hall, Zvonimir Rakamaric; Organization:University of Utah;NSF Organization:CCF Start Date:08/01/2017; Award Amount:\$1,200,000.00; July 31, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1704715&HistoricalAwards=false

The project implements **methods to improve the resource-efficiency of numerical computations on a variety of computing machines**, ranging from supercomputers to mobile devices, by adaptively reducing data-precision based on the needs of the application. Efficiently adapting data-precision permits these machines to run larger computations and also improves the overall performance (including energy consumption) made possible by a combination of reduced computational burden as well as reduced data movement. The intellectual merits of this project are to research and develop the key steps to understand the nature of applications and computing systems, and to suitably **minimize computing demands through targeted data-precision adjustment**. Broader impacts of the work include training graduate students and releasing tools that the community can employ in future hardware and software product, to help minimize overall energy consumption and improve performance.

Unused precision in floating-point computations ends up wasting allocated space in caches, and also causes unnecessary data movement. The technical parts of the project are to identify as well as pursue opportunities for optimally allocating precision, and to efficiently implement such allocation methods in actual codes. In addition to developing new algorithms to tune precision assisted by automated learning methods, the project develops symbolic analysis methods to serve novel roles in floating-point instruction selection and optimization in the form of superoptimizers. These tools will be released to a community of researchers interested in working toward exascale computing, and deploying applications in safety-critical devices. This work represents a synergistic combination of the investigator's skills ranging through high performance computing, formal methods, and compiler technologies.

[NSF HPC 예비후보07]

Enhancing the Open MPI Software for Next Generation Architectures and Applications

Collaborative Research: SI2-SSI: EVOLVE:

Award Number:1664142; Principal Investigator:George Bosilca; Co-Principal Investigator:Aurelien Bouteiller, Thomas Herault; Organization:University of Tennessee Knoxville;NSF Organization:OAC Start Date:06/01/2017; Award Amount:\$1,566,215.00; May 31, 2021 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1664142&HistoricalAwards=false

For nearly two decades, the **Message Passing Interface (MPI)** has been an essential part of the High-Performance Computing ecosystem and consequently a key enabler for important scientific breakthroughs. It is a fundamental building block for most large-scale simulations from physics, chemistry, biology, material sciences as engineering. **Open MPI** is an open source implementation of the MPI specification, widely used and adopted by the research community as well as industry. The Open MPI library is jointly developed and maintained by a consortium of academic institutions, national labs and industrial partners. It is installed on virtually all large-scale computer systems in the US as well as in the rest of the world. The goal of this project is to enhance and modernize the Open MPI library in the context of the ongoing evolution of modern computer systems, and to ensure its future operability on all upcoming architectures. We aim at implementing fundamental software techniques that can be used in many-core systems to execute MPI-based parallel applications more efficiently, and to tolerate process and memory failures at all scales, from current systems, up to the extreme scales expected before the end of the decade.

Open MPI is an open source implementation of the Message Passing Interface (MPI) specification. The MPI API is currently being extended to consider the needs of application developers in terms of efficiency, productivity and resilience. The project will also support academic involvement in the design, development and evaluation of the Open MPI software, and ensure academic presence in the MPI Forum. The goal of this proposal is **to enhance the Open MPI software library**, focusing on two aspects: (1) **Extend Open MPI to support new features** of the MPI specification. Open MPI will continue to support all new features of current and upcoming MPI specifications. The two most significant areas within the context of this proposal are (a) extensions to better support hybrid programming models and (b) support for fault tolerance in MPI applications. To improve support for hybrid programming models, the MPI Forum is currently considering introducing the notion of MPI Endpoints, which could be used by different threads of an MPI rank to instantiate multiple separate communication contexts. The goal within this project is to develop an implementation of endpoints to support effective hybrid programming model, and to extend the concept to other aspects of parallel applications such as File I/O operations. One of the project partners (UTK) leads the current proposal in the MPI Forum to expose failures and ensure the continuation of the execution of MPI applications. In the context of this SSI proposal, the goal is to harden, improve, and expand the support of the existing ULFM implementation in Open MPI and thus enable end-users to design application-specific resilience approaches for future platforms. (2) **Enhance the Open MPI core to support new architectures and improve scalability.** While Open MPI has demonstrated very good scalability in the past, there is significant work to be done to ensure similarly good performance on future architectures. Specifically, we propose a groundbreaking rework of the startup environment that will improve process launch scalability, increase support for asynchronous progress of operations, enable support for accelerators, and reduce sensitivity to system noise. The project would also enhance the support for File I/O operations as part of the Open MPI package by expanding our work on highly scalable collective I/O operations through delegation and exploring the utilization of burst buffers as temporary storage.

[NSF HPC 예비후보08]

Ubiquitous Access to Transient Data and Preliminary Results via the SeedMe Platform

CIF21 DIBBs:

Award Number:1443083; Principal Investigator:Amit Chourasia; Co-Principal Investigator:Michael Norman; Organization:University of California-San Diego;NSF Organization:OAC Start Date:10/01/2014; Award Amount:\$1,329,379.00; Relevance:64.16; September 30, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1443083&HistoricalAwards=false

Computational simulations have become an indispensable tool in a wide variety of science and engineering investigations. Quick and effective assessments of the resulting data are necessary for efficient use of researcher time and computation resources, but this process is complicated when a large collaborating team is geographically dispersed and/or some team members do not have direct access to the computation resource and output data. Current methods for sharing and assessing transient data and preliminary results are cumbersome and labor intensive; each research team must create their own scripts and ad hoc procedures to push data from system to system and user to user. Better tools and cyberinfrastructure are needed to support preliminary results sharing for collaborating computational science teams.

This project develops **web-based building blocks and cyberinfrastructure** to enable **easy sharing and streaming of transient data and preliminary results** from computing resources to a variety of platforms, from mobile devices to workstations, making it possible to quickly and conveniently view and assess results and provide an essential missing component in High Performance Computing and cloud computing infrastructure.

Amit Chourasia, Mona Wong, Dmitry Mishin, David R. Nadeau, and Michael Norman. "SeedMe: A scientific data sharing and collaboration platform," *Juried Conference Paper*, 2016.
[doi:10.1145/2949550.2949590](https://doi.org/10.1145/2949550.2949590)

[NSF HPC 예비후보09]

Sustainable Open-Source Quantum Dynamics and Spectroscopy Software

SI2-SSI:

Award Number:1663636; Principal Investigator:Xiaosong Li; Co-Principal Investigator:Albert DePrince, Eitan Geva, Anne McCoy; Organization:University of Washington;NSF Organization:OAC Start Date:09/01/2017; Award Amount:\$1,500,000.00; August 31, 2021 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1663636&HistoricalAwards=false

Advances in experimental techniques that use multiple light sources to probe chemical systems provide unprecedented ability to understand the evolution of important and interesting chemical processes. Theory and simulation are essential to extract the maximum amount of information from these experiments. Thus, there is a strong need for **user-friendly computer programs** to model these processes and to provide physical insights to experimentalists. To this end, this project is developing **an innovative software package (ChronusQ)** capable of modeling many types of experimental measurements that involve matter interacting with multiple incident light sources. The physical insights gleaned through application of ChronusQ will be useful for the advancement of renewable energy and information technologies. In addition, given the nature of the project, students and postdocs will have unique opportunities to gain experience in high performance computing software development through direct collaborations with engineers from industrial partnerships

The overarching goal of the proposed activity is **to develop an innovative software platform, namely Chronus Quantum (ChronusQ)**, which is capable of modeling different types of time resolved multidimensional spectral signals using quantum electronic and nuclear dynamics. ChronusQ performs quantum dynamic simulations of the same light and matter interactions that occur in time resolved multidimensional spectroscopies directly in the time domain. The time correlated experimental observables required to model multidimensional spectra can then be extracted from these simulations. By providing a time dependent, state specific interpretation for the chemical dynamics encoded in multidimensional spectra, the proposed development will aid in the design of new molecules and materials that exhibit the desirable optical characteristics, with the potential for transformative impact in the broader scientific community and beyond. The ChronusQ software represents the next frontier for innovations in **computational spectroscopy**, which will have a far reaching impact on education and research in multidisciplinary scientific communities including chemistry, physics, nanoscience and surface science, and other fields relying on these cutting edge spectroscopic methods.

[NSF HPC 예비후보10]

High Performance and Scalable Fabric Analysis, Monitoring and Introspection Infrastructure for HPC and Big Data

SI2-SSI: FAMI:

Award Number:1664137; Principal Investigator:Dhabaleswar Panda; Co-Principal Investigator:Kevin Manalo, Karen Tomko, Xiaoyi Lu, Hari Subramoni; Organization:Ohio State University;NSF Organization:OAC Start Date:07/01/2017; Award Amount:\$800,000.00; Relevance:50.8; June 30, 2020 (Estimated)

https://nsf.gov/awardsearch/showAward?AWD_ID=1664137&HistoricalAwards=false

As the computing, networking, heterogeneous hardware, and storage technologies continue to evolve in High-End Computing (HEC) platforms, it becomes increasingly essential and challenging to understand the interactions between time-critical High-Performance Computing (HPC) and Big Data applications, the software infrastructures upon which they rely for achieving high-performing portable solutions, the underlying communication fabric these high-performance middlewares depend on and the schedulers that manage HPC clusters. Such understanding will enable all involved parties (application developers/users, system administrators, and middleware developers) to maximize the efficiency and performance of the individual components that comprise a modern HPC system and solve different grand challenge problems. There is a clear need and unfortunate lack of a high-performance and scalable tool that is capable of analyzing and correlating the communication on the fabric with the behavior of HPC/Big Data applications, underlying middleware and the job scheduler on existing large HPC systems. The proposed synergistic and collaborative effort, undertaken by a team of computer and computational scientists from OSU and OSC, aims **to create an integrated software infrastructure for high-performance and scalable Fabric Analysis, Monitoring and Introspection for HPC and Big Data**. This tool will achieve the following objectives: 1) be portable, easy to use and easy to understand, 2) have high performance and scalable rendering and storage techniques and, 3) be applicable to the different communication fabrics and programming models that are likely to be used on existing large HPC systems and emerging exascale systems. The transformative impact of the proposed research and development effort is to design a comprehensive analysis and performance monitoring tool for applications of current and next generation multi petascale/exascale systems to harness the maximum performance and scalability.

The proposed research and the associated infrastructure will have a significant impact on enabling optimizations of HPC and Big Data applications that have previously been difficult to provide. These potential outcomes will be demonstrated by using the proposed framework to validate a variety of HPC and Big Data benchmarks and applications under multiple scenarios. The **integrated middleware and tools** will be made publicly available to the community through public repositories and publications in the top forums, enabling other MPI and Big Data stacks to adopt the designs. Research results will also be disseminated to the collaborating organizations of the investigators to impact their HPC software products and applications. The proposed research directions and their solutions will be used in the curriculum of the PIs to train undergraduate and graduate students, including under-represented minorities and female students. The technical challenges addressed by the proposal include: 1) Scalable visualization of large and complex HEC networks so as to provide a near instant rendering to end users, 2) A generalized data gathering scheme which is easily portable to multiple communication fabrics, novel compute architectures and high-performance middleware, 3) Enhanced data storage performance through optimized database schemas and the use of memory-backed key value stores/databases, 4) Support in MPI, PGAS, and Big Data libraries to enable the proposed monitoring, analysis, and introspection framework, and 5) Enabling deeper introspection of particular regions of application. The research will also be driven by a set of HPC and Big Data applications. The transformative impact of the proposed research and development effort is to design a comprehensive analysis and performance monitoring tool for applications of current and next generation multi petascale/exascale systems to harness the maximum performance and scalability.

<부록 15> NSF 클라우드 예비 후보

[NSF 클라우드 예비후보01]

Community Infrastructure To Expand the Frontiers of Cloud Computing Research

CloudLab Phase II:

Award Number:1743363; Principal Investigator:Robert Ricci; Co-Principal Investigator:Glenn Ricart, Michael Zink, Kuang-Ching Wang, Srinivasa Akella; Organization:University of Utah;NSF Organization:CNS Start Date:10/01/2017; Award Amount:\$5,833,587.00; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1743363&HistoricalAwards=false

This project seeks to **extend** the capabilities and capacity of the **Cloudlab** compute cloud in the **second phase** of the **NSFCloud program**. Since CloudLab began service in mid-2015, this **experimental research infrastructure** has become a critical experimental resource for the computer science research community. CloudLab has served over 2,000 users in over 39,000 experiments spanning more than 400 projects, with experimenters coming from nearly every state. The Cloudlab project continues to increase its support of computing for the domain science research communities, and support of technology transfers to commercial cloud technology providers and users.

The proposed Phase II activity seeks to expand the testbed capacity and capabilities through 11 enumerated hardware and software extensions to satisfy the growing needs of leading computer science systems researchers. The three major areas of new technology investments are in 1) new cloud networking technology, 2) support for new cloud architectures, and 3) increased programmability in cloud network data planes.



<https://www.cloudlab.us>

Flexible, scientific infrastructure for research on the future of cloud computing. Researchers use CloudLab to build their own clouds, experimenting with new architectures that will form the basis for the next generation of computing platforms.

The CloudLab Team

CloudLab is a project of the University of Utah, Clemson University, the University of Wisconsin Madison, the University of Massachusetts Amherst, Raytheon BBN Technologies, and US Ignite.

We've built widely-used testbeds for the computer science research community for decades, including Emulab, parts of GENI, and Apt. We're also heavily involved in reaching out to the community of computational research through the ACI-REF initiative, next-generation applications with transformative public benefit through US-IGNITE, and through a partnership with HTCondor, CloudLab users will be able to "Opt-in" to compute jobs as real workloads if they wish.

To design and build the CloudLab facility, we're partnering with three vendors: Cisco, Dell, and HP. Seagate has also provided a generous donation of hard drives.

CloudLab is part of the National Science Foundation's NSFCloud program.

[NSF 클라우드 예비후보02]

Chameleon: A Large-Scale, Reconfigurable Experimental Environment for Cloud Research

Collaborative Research:

https://www.chameleoncloud.org/media/cms_page_media/17/Chameleon-fire-geni_OSpyKvd.pdf

Award Number:1743358; Principal Investigator:Katarzyna Keahey; Co-Principal Investigator;; Organization:University of Chicago;NSF Organization:CNS Start Date:10/01/2017; Award Amount:\$3,102,183.00; eptember 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1743358

Award Number:1743354; Principal Investigator:Daniel Stanzione; Co-Principal Investigator;; Organization:University of Texas at Austin;NSF Organization:CNS Start Date:10/01/2017; Award Amount:\$2,367,716.00; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1743354&HistoricalAwards=false

Award Number:1743324; Principal Investigator:Joel Mambretti; Co-Principal Investigator;; Organization:Northwestern University;NSF Organization:CNS Start Date:10/01/2017; Award Amount:\$100,000.00; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1743324&HistoricalAwards=false

One of **two** NSFCloud mid-scale experimental computing research infrastructures, **Chameleon** is a **deeply reconfigurable cloud computing testbed** supporting large-scale Computer Science experimentation. Located at the University of Chicago and the Texas Advanced Computing Center, it is composed of over 600 compute nodes and a total of **5 petabytes of storage** supporting data-intensive science. Since its public availability in July 2015, Chameleon has attracted a community of over 1,300 users working on over 200 education and research projects.

In Phase 2 of the NSFCloud program, this project will expand Chameleon capabilities and capacity, and will fulfill its mission of serving as a testbed that is used extensively by the research community by:

- 1) broadening the set of supported experiments through the addition of new hardware and software supporting high bandwidth and Software Defined Networking (SDN)
- 2) providing operational sustainability by streamlining operations, packaging the testbed as open source software that can be leveraged by others, and developing a process for adding resources to the testbed
- 3) provide support for managing, understanding, comparing and reproducing Computer Science experiments, and
- 4) focusing effort on building an engaged community around the testbed via both traditional outreach including targeted training and education efforts, and direct personalized outreach to experimenters.



The Chameleon architecture consists of a set of standard cloud units (SCUs), each of which is a single rack with 42 compute nodes, 4 storage nodes attached to 128TB of local storage in configurable arrays, and an OpenFlow compliant network switch. In addition to the homogeneous SCUs, a variety of heterogeneous hardware types is available to experiment with alternative technologies. The testbed also includes a shared infrastructure with a persistent storage system accessible across the testbed, a top-level network gateway to allow access to public networks, and a set of management and provisioning servers to support user access, control, monitoring and configuration of the testbed. Chameleon is physically distributed between the Texas Advanced Computing Center (TACC) and the University of Chicago (UC) through 100Gbps Internet2 links, to allow users to examine the effects of a distributed cloud.

[NSF 클라우드 예비후보03]

Self-Monitoring Virtual Machines for Performance Guarantees in Public Clouds

CSR: Small:

Award Number:1718084; Principal Investigator:Yinqian Zhang; Co-Principal Investigator;; Organization:Ohio State University;NSF Organization:CNS Start Date:10/01/2017; Award Amount:\$500,000.00; Relevance:87.71; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1718084&HistoricalAwards=false

With its massive pooling and multiplexing of computing resources, the cloud offers both large organizations and small businesses the prospect of lower information technology costs, lighter administrative burdens, and rapid scaling of resources. However, multi-tenancy in public clouds makes critical computing resources, such as processor, memory, I/O devices, and storage, shared among virtual machines that are operated by different users. The performance of applications running in public clouds, therefore, may be affected by their neighboring virtual machines due to contention on the shared computing resources. Nonetheless, existing cloud performance monitoring tools do not offer visibility into hardware resources; cloud users have no choice but to blindly run computations on these public services in the hope that the performance is not negatively affected, for instance, by the neighbor's resource-depleting applications. The lack of performance guarantees is a hurdle faced by all cloud users to fully embrace the economic benefit of cloud computing, and especially by those whose applications demand stability and predictability of the runtime environments. This proposed project aims to solve this problem **by developing novel techniques that allow cloud users to monitor the resource contention on the physical cloud servers without the help of the cloud providers.** Specifically, the proposed work entails the design, implementation, and evaluation of self-monitoring virtual machines, which leverage the nested virtualization technology and side-channel analysis techniques to monitor the contention in shared computing resources and proactively migrate nested virtual machines to avoid severe performance degradation.

The project will make broader societal impacts in the following aspects: First, new education tools will be developed through the proposed project. Specifically, one of the outcomes of the intended research will be Amazon machine images with which a derivative cloud can be created on top of public clouds. Enabled by this derivative cloud, students of the operating systems or system security courses can obtain hands-on experience with cloud computing; they will also have access to nested virtualization environments to conduct operating system kernel development. Second, the project will be integrated into NSF's LSAMP (Louis Stokes Alliances for Minority Participation) program, to help increase underrepresented minority student recruitment, retention, and attainment of STEM degrees, and also to enhance the participation of underrepresented minority students in system research. Third, the project will produce open-source tools that enable self-monitoring VMs in public clouds, which will be made available in the form of source code (available on the project homepage) and Amazon machine images, to encourage adaptation and adoption of the developed techniques.

[NSF 클라우드 예비후보04]

New Abstractions for First-hop Networking in Cloud Data Centers

NeTS: Small:

Award Number:1717039 ; Principal Investigator:Srinivasa Akella; Co-Principal Investigator:Michael Swift; Organization:University of Wisconsin-Madison;NSF Organization:CNS Start Date:09/01/2017; Award Amount:\$500,000.00; Relevance:87.71; August 31, 2020 (Estimated) August 31, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1717039&HistoricalAwards=false

This project seeks to improve the performance of future data center compute servers and networks, enabling a growing number of performance-sensitive applications to co-exist on each server. To allow multiple applications on a server to drive the Network Interface Card (NIC) at high speed, this research aims to rethink the mechanisms employed in various layers of the server networking stack, and create new interfaces and abstractions between the layers and for applications. This research will lead to improved scalability, utilization, and efficiency of cloud computing infrastructure, allowing providers to support many more applications and tenants while maintaining their current infrastructure footprint.

This project seeks **to develop new algorithms and programming abstractions for server operating systems and Network Interface Cards (NICs)** to improve scheduling of outbound network traffic. These algorithms will overcome current performance bottlenecks, and will improve the ability for multiple applications to drive a NIC at high speed, with the ambitious goal of sustaining line rates while meeting application latency/throughput/fairness objectives as well as infrastructure-wide utilization and efficiency objectives. The PIs plan to integrate the proposed research into courses at UW-Madison. The PIs will organize various 'network programming bootcamps' aimed at network engineers, undergraduate students, and high school students from under-represented groups.

[NSF 클라우드 예비후보05]

Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

CC*DNI DIBBs:

Award Number:1541215; Principal Investigator:David Lifka; Co-Principal Investigator:Thomas Furlani, Richard Wolski; Organization: Cornell University; NSF Organization:OAC Start Date:10/01/2015; Award Amount:\$6,114,695.00; Relevance:83.97; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1541215&HistoricalAwards=false

The ability to aggregate, share, and analyze important large data sets while optimizing time-to-science is essential to support multi-disciplinary and multi-institutional data-driven discovery. This project is deploying a **federated cloud computing system** in **New York State** and **California** comprised of data infrastructure building blocks designed to support scientists requiring flexible workflows and analysis tools for large-scale data sets. Data challenges from seven different communities-earth and atmospheric sciences, finance, chemistry, astronomy, civil engineering, genomics, and food science-are being addressed using a rich set of open source software, optimized frameworks, and cloud usage modalities. The federated cloud is operating at Cornell University (project lead) and at partner sites at the University at Buffalo and the University of California, Santa Barbara. The project team is supporting multi-disciplinary research groups with over forty global collaborators and documenting science use cases. The broader goal of this project is to develop a federated cloud model that encourages and rewards institutions for sharing large-scale data analysis resources that can be expanded internally with common, incremental building blocks and externally through meaningful collaborations with other institutions, public clouds, and NSF cloud resources.

Project documentation and webinars feature best practices and include how to create Virtual Machine instances, run at federated sites, burst to Amazon Web Services, and access, move, and store large-scale data. A new tool for cloud metrics is being built into **Open XDMoD (XD Net Metrics on Demand)** that features QBETS (Queue Bounds Estimation from Time Series) statistics to enable users to make online forecasts of future performance and allocation level availability as well as to predict when to burst from federation resources. A new allocations and accounting model allows institutional administrators to track utilization across federated sites and use this data as an exchange mechanism. These tools provide a better understanding of how the sharing of data infrastructure building block capacity across institutional boundaries can create wider science and engineering collaborations and increase data sharing in a scalable and sustainable way.

Andy Rosales Elias, Nevena Golubovic, Chandra Krintz, and Rich Wolski. "Where's the Bear? ? Automating wildlife image processing using IoT and edge cloud systems.," *Juried Conference Paper*, 2017, p. 247. [doi:10.1145/3038912.3052649](https://doi.org/10.1145/3038912.3052649)

Hiranya Jayathilaka, Chandra Krintz, and Rich Wolski. "Performance monitoring and root cause analysis for cloud-hosted web applications," *Juried Conference Paper*, 2017, p. 469. [doi:10.1145/3038912.3052649](https://doi.org/10.1145/3038912.3052649)

Nana Y.D. Ankrah, Junbo Luan, and Angela E. Douglas. "Cooperative metabolism in a three-partner insect-bacterial symbiosis revealed by metabolic modeling," *Journal*, v.199, 2017, p. e00872-16. [doi:10.1128/JB.00872-16](https://doi.org/10.1128/JB.00872-16)

Wolski, R. and Brevik, J.. "Providing Statistical Reliability Guarantees in the AWS Spot Tier.," *24th High Performance Computing Symposium (HPC 2016)*, 2016, p. 13. [doi:Society for Computer Simulation International, 2016](https://doi.org/10.1145/2898888.2898888)

[NSF 클라우드 예비후보06]

An Architecture for Scalable Verifiable Computing

TWC: Medium: Apollo:

Award Number:1514261; Principal Investigator:Charalampos Papamanthou; Co-Principal Investigator:Jonathan Katz, Amol Deshpande, Elaine Shi; Organization:University of Maryland College Park;NSF Organization:CNS Start Date:07/01/2015; Award Amount:\$1,162,868.00; Relevance:80.98; June 30, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1514261&HistoricalAwards=false

Cloud computing enables computationally limited parties to outsource the storage of, and computation on, massive amounts of data. Such outsourced computations, however, might be performed incorrectly, due to implementation errors or malicious behavior. **Protocols for verifiable computation** allow an untrusted server performing such computations to also provide succinct proofs that the returned results are correct. Drawing on expertise from cryptography, programming languages, databases, and systems, this project develops **a novel architecture (Apollo) for automatically producing scalable verifiable-computation protocols**. Apollo has the potential to speed up verifiable computations by orders of magnitude while at the same time scaling to large datasets.

This research focuses on **cryptographic "gadgets"** suited for **verifiable computation** in different settings, as well as programming languages and type systems for easily expressing verifiable computation in terms of these lower-level gadgets. A major goal of the project is to develop **a working prototype** of a **verifiable SQL database**. Results of this research will be disseminated to a wide audience through publications and open-source implementations, and will be integrated into a graduate course on cloud computing security at the University of Maryland.

C. Christopher Erway, Alptekin Küpçu, Charalampos Papamanthou, and Roberto Tamassia. "Dynamic Provable Data Possession," *ACM Trans. Inf. Syst. Secur.*, v.17, 2015, p. 15.

Charalampos Papamanthou, Roberto Tamassia, Nikos Triandopoulos. "Authenticated Hash Tables Based on Cryptographic Accumulators," *Algorithmica*, v.74, 2016, p. 664.

Eli Ben-Sasson, Iddo Bentov, Alessandro Chiesa, Ariel Gabizon, Daniel Genkin, Matan Hamilis, Evgenya Pergament, Michael Riabzev, Mark Silberstein, Eran Tromer, Madars Virza. "Computational Integrity with a Public Random String from Quasi-Linear PCPs," *EUROCRYPT*, 2017.

Yupeng Zhang, Daniel Genkin, Jonathan Katz, Dimitrios Papadopoulos, Charalampos Papamanthou. "vSQL: Verifying General SQL Queries over Dynamic Outsourced Databases," *SSP (IEEE Symposium on Security and Privacy)*, 2017.

Yupeng Zhang, Jonathan Katz, Charalampos Papamanthou. "An Expressive (Zero-Knowledge) Set Accumulator," *EUROSSP*, 2017.

[NSF 클라우드 예비후보07]

Language-Hardware Co-Design for Practical and Verifiable Information Flow Control

TWC: Medium:

Award Number:1513797; Principal Investigator:Gookwon Suh; Co-Principal Investigator:Andrew Myers; Organization:Cornell University;NSF Organization:CNS Start Date:07/01/2015; Award Amount:\$1,200,000.00; Relevance:79.32; June 30, 2019 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1513797&HistoricalAwards=false

Current cloud computing platforms, mobile computing devices, and embedded devices all have the security weakness that they permit information flows that violate the confidentiality or integrity of information. This project explores an integrated approach in which **software and hardware are co-designed with** strong, comprehensive, verifiable security assurance. The goal is **to develop a methodology for designing systems in which all forms of information flow are tracked, at both the hardware and software levels, and between these levels.** This can have a significant impact on how computing systems are designed, and could make the next generation of computing devices and platforms inherently more secure.

The project investigates a novel hardware description language, that allows information flow and timing channels to be soundly tracked at the hardware level, while permitting fine-grained sharing and reuse of hardware resources across multiple security levels. The practicality of this approach will be demonstrated on a high-performance microprocessor with verified-secure information flow which is designed as part of the project. The project explores novel techniques for identifying information flows that must be communicated between the software and hardware levels in order for the hardware to enforce security, and operating-system mechanisms for automatically virtualizing rich language-level security policies onto the more limited palette of security levels representable in hardware. The project uses rigorous verification of the new language-level and hardware-level techniques to demonstrate that applications and hardware developed according to the new methodology provably enforce secure information flow throughout the computing system.

Andrew Ferraiuolo, Yao Wang, Danfeng Zhang, Andrew C. Myers, and G. Edward Suh. "Lattice Priority Scheduling: Low-Overhead Timing Channel Protection for a Shared Memory Controller," *Proceedings of the 22nd International Symposium on High-Performance Computer Architecture (HPCA)*, 2016.

Yao Wang, Andrew Ferraiuolo, Danfeng Zhang, Andrew C. Myers, and G. Edward Suh. "SecDCP: Secure Dynamic Cache Partitioning for Efficient Timing Channel Protection," *Proceedings of the the 53rd Design Automation Conference (DAC)*, 2016.

Yao Wang, Benjamin Wu, and G. Edward Suh. "Secure Dynamic Memory Scheduling Against Timing Channel Attacks," *The 23rd International Symposium on High-Performance Computer Architecture*, 2017.

Andrew Ferraiuolo, Rui Xu, Danfeng Zhang, Andrew C. Myers, and G. Edward Suh. "Verification of a Practical Hardware Security Architecture Through Static Information Flow Analysis," *The 22nd International Conference on Architectural Support for Programming Languages and Operating Systems*, 2017.

Isaac Sheff, Tom Magrino, Jed Liu, Andrew C. Myers, and Robbert Van Renesse. "Safe serializable secure scheduling: transactions and the trade-off between security and consistency," *ACM Conference on Computer and Communication Security (CCS)*, 2016, p. 229. [doi:978-1-4503-4139-4](https://doi.org/10.1145/2976764.2976804)

Yizhou Zhang, Guido Salvaneschi, Quinn Beightol, Barbara Liskov, and Andrew C. Myers. "Accepting blame for safe tunneled exceptions," *37th ACM SIGPLAN Conf. on Programming Language Design and Implementation (PLDI)*, 2016, p. 281. [doi:978-1-4503-4261-2](https://doi.org/10.1145/2976764.2976804)

[NSF 클라우드 예비후보08]

Jetstream - A Self-Provisioned, Scalable Science and Engineering Cloud Environment

High Performance Computing System Acquisition:

Award Number:1445604; Principal Investigator:Craig Stewart; Co-Principal Investigator:Ian Foster, Matthew Vaughn, Nirav Merchant, James Taylor; Organization:Indiana University;NSF Organization:OAC Start Date:12/01/2014; Award Amount:\$12,340,710.00; Relevance:87.72; November 30, 2019 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1445604&HistoricalAwards=false

High Performance Computing System Acquisition: Jetstream - a self-provisioned, scalable science and engineering cloud environment

Jetstream will be a new type of computational research resource open for the national (nonclassified) research community - a data analysis and computational resource that US scientists and engineers will use interactively to conduct their research anytime, anywhere. Jetstream will complement current NSF-funded computational resources and bring a cloud-based system to the NSF computational resources incorporating the best elements of commercial cloud computing resources with some of the best software in existence for solving important scientific problems. This system will enable many US researchers and engineers to make new discoveries that are important to understanding the world around us and will help researchers make new discoveries that improve the quality of life of American citizens.

In terms of technical details, Jetstream will be a configurable large-scale computing resource that leverages both on-demand and persistent virtual machine technology to support a much wider array of software environments and services than current NSF resources can accommodate. As a fully configurable "cloud" resource, Jetstream bridges the obvious major gap in the current ecosystem, which has machines targeted at large-scale High-Performance Computing, high memory, large data, high-throughput, and visualization resources. As the open cloud for science, Jetstream will:

*Provide "self-serve" academic cloud services, enabling researchers or students to select a VM image from a published library, or alternatively to create or customize their own virtual environment for discipline- or task-specific personalized research computing.

*Host persistent VMs to provide services beyond the command line interface for science gateways and other science services. For example, Jetstream will become a primary host of the popular Galaxy scientific workbench and its main datasets, bringing many Galaxy users to the NSF ecosystem from day one.

Craig A. Stewart, David Y. Hancock, Matthew Vaughn, Lee Liming, Nirav Merchant, Therese Miller, John M. Lowe, Dan Stanzione, James Taylor, Edwin Skidmore. "Jetstream - Performance, Early Experiences, and Early Results," *Proceedings of the XSEDE16 Conference. St. Louis, MO.*, 2016, p. 1. [doi:10.1145/2792745.2792774](https://doi.org/10.1145/2792745.2792774)

Eric Coulter, Jeremy Fisher, Barbara Hallock, Richard Knepper, Dave Lifka, JP Navarro, Marlon Pierce, Craig Stewart. "Cyberinfrastructure as a Platform to Facilitate Effective Collaboration between Institutions and Support Collaboratories," *Proceedings of the 2016 ACM on SIGUCCS Annual Conference*, 2016, p. 37. [doi:10.1145/2974927.2974962](https://doi.org/10.1145/2974927.2974962)

Craig A. Stewart. "Preserving Scientific Software . . . in a Usable Form?," *EDUCAUSE Review*, v.51, 2016.

Milad Memarzadeh, Carl Boettiger. "Measurement uncertainty matters: ecological management using POMDPs," *Proceedings of the National Academy of Sciences*, 2017. [doi:10.1101/055319](https://doi.org/10.1101/055319)

Knepper, Richard and Coulter, Eric and Pierce, Marlon and Marru, Suresh and Pamidighantam, Sudhakar. "{Using the Jetstream Research Cloud to provide Science Gateway resources}," *2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, 2017, p. 753--757. [doi:10.1109/CCGRID.2017.121](https://doi.org/10.1109/CCGRID.2017.121)

[NSF 클라우드 예비후보09]

Enabling Flexible Middlebox Processing in the Cloud

NeTS: Medium: Collaborative Research:

Award Number:1440056; Principal Investigator:Vyas Sekar; Co-Principal Investigator;;
Organization:Carnegie-Mellon University;NSF Organization:CNS Start Date:01/01/2014; Award
Amount:\$500,000.00; Relevance:87.71; August 31, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1440056&HistoricalAwards=false

Enterprises rely on **specialized network appliances** or **middleboxes** such as load balancers, intrusion detection and prevention systems, and WAN optimizers in order to meet critical performance optimization, security, and policy compliance requirements. With the advent of cloud computing, such middlebox processing will play an increasingly critical role in cloud deployments due to two key factors: 1) As enterprises move their IT infrastructure to the cloud, they want to leverage the same performance and security benefits for applications running in the cloud; and 2) Enterprises want to reduce their infrastructure and management costs by offloading middlebox functionality to cloud providers to leverage the elastic scaling and migration benefits offered by cloud computing.

Unfortunately, cloud customers and providers today lack the necessary abstractions and mechanisms for enabling this transition. At a high-level, the problem is that these workloads are drastically different from traditional computation and storage services for which cloud computing has been extremely successful. This raises fundamental challenges along several dimensions: the need for flexible composition or chaining of network services; the increased impact of network-level performance on such workloads; the inherent difficulty in identifying bottlenecked resources in multiplexed cloud deployments; and the inability to reason about correct and consistent operation of stateful network processing in dynamic deployment scenarios.

This project will bridge this disconnect by addressing foundational issues in the design and implementation of (1) policy frameworks, elastic scaling algorithms, and software-defined controllers for enterprise administrators to translate their requirements into an actual physical realization; (2) algorithms for intelligent network-level placement, traffic engineering, and topology design for cloud providers to support such workloads; and (3) new abstractions for managing and manipulating the middlebox-associated state of the network.

Broader Impact: This work will inform the critical industry evolution as enterprises and cloud providers are attempting to realize the benefits of ?network virtualization?. Furthermore, the project will enable new dimensions of flexibility for network deployments that do not exist today---democratizing the benefits of middleboxes to small businesses; providing the ability to elastically scale network-level services to meet application demands; and enabling live migration of entire enterprise deployments across physical infrastructures. The project will generate new course materials on software-defined networking and cloud computing and tightly integrate research with education to help students become experts in these emerging domains. The software tools and benchmark measurement data produced by the research will inform the industry transition and future academic work on such middleboxes-in-the-cloud deployments. Finally, while the project focuses on middleboxes in cloud deployments, the technical foundations developed therein will apply to traditional enterprise and ISP networks as well.

[NSF 클라우드 예비후보10]

A Computing Cloud for Graphical Simulation

CSR: Medium:

Award Number:1409847; Principal Investigator:Philip Levis; Co-Principal Investigator:패키지인
PhysBamr:Ronald Fedkiw; Organization:Stanford University;NSF Organization:CNS Start
Date:08/01/2014; Award Amount:\$1,093,026.00; Relevance:87.71; July 31, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1409847&HistoricalAwards=false

high-performance, high-cost nodes. This research aims to answer the question -- is it possible to run graphical simulations in the computational cloud? -- by designing and implementing Nimbus, **a software for graphical simulation in the computing cloud**. The goal is to be able to run large, complex simulations using on-demand cloud computing systems. **Nimbus** supports **PhysBAM, an open-source graphical simulation package** developed and maintained by Principal Investigator Fedkiw. The project will collaborate with existing PhysBAM users to support the Nimbus software for broader use and adoption.

Nimbus focuses on three important principles to support graphical simulations running on hundreds to thousands of cloud servers. First is decoupling data access and layout. Nimbus represents data in three layers: program, logical, and physical. These layers separate the units which a program operates on (program) from the units which the Nimbus software manages and transfers (logical) from how they are laid out in actual computer memory (physical). Second is non-uniform, geometry-aware data placement. Nimbus uses the fact that simulations have a basic underlying geometry to intelligently place data and computation. This geometry is explicit in the Nimbus software, which knows that nearby regions of the simulation should be placed on nearby computers. Third is dynamic assignment and load balancing: Graphical simulations today divide the simulation volume equally across computers, despite the fact that some regions require much more computation than others. Nimbus divides a simulation into a larger number of smaller partitions, which it dynamically assigns and moves as load changes to reduce running time while considering inter-partition communication. These three principles allow Nimbus to provide tremendous flexibility. The system breaks a simulation into small pieces that a controller computer sends to worker computers to compute. These worker computers decide when to schedule these simulation pieces and how to assign processors to different pieces. The runtime automatically moves data in the most efficient manner possible as needed, compressing data and replicating it when having multiple copies for different pieces increases performance. Discovering how these applications can be run on modern data center computing systems will help bring arithmetically intensive scientific computing to the cloud. As Exascale and other supercomputing efforts gain momentum, their scale will need to deal with the same issues cloud systems have been tackling for the past decade, stragglers, failures, and heterogeneity. By focusing on one particular compelling application, this work will establish an intellectual framework for future, broader efforts

[NSF 클라우드 예비후보11]

NSF Net-centric and Cloud Software and Systems

I/UCRC:

Award Number:1361806; Principal Investigator:Krishna Kavi; Co-Principal Investigator:Robert Akl; Organization:University of North Texas;NSF Organization:CNS Start Date:04/15/2014; Award Amount:\$1,092,794.00; March 31, 2019 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1361806&HistoricalAwards=false

The **NSF Net-centric and Cloud Software and Systems Industry/University Cooperative Research Center (NCSS I/UCRC)** is a multi-university cooperative established in **2009** with a collaborative research partnership between the University of North Texas (UNT) and the University of Texas at Dallas (UTD), with Southern Methodist University (SMU) as an affiliated academic site. The original academic sites (UNT and UTD, with SMU as an affiliated member) will complete **Phase I in February 2014** and are seeking to continue operations as Phase II I/UCRC sites. The research focus of the NCSS I/UCRC is on cost-effective methods of enabling the rapid design and deployment of highly efficient and dependable net centric and cloud hardware/software systems that must operate in various environments subject to potentially stringent response time, power, heat, security and resource constraints. By combining the diverse capabilities and expertise of the participating academic institutions and the real-world experiences of engineers from well-known innovative high tech companies, the center will enhance the research capabilities of all the participants. The NCSS I/UCRC will develop innovative and practical technologies for modeling, analysis, design, implementation, verification and validation, deployment, and evolution of a variety of net-centric and cloud systems. This systems-oriented research is aimed at enabling the coordinated hardware/software development of highly dependable integrated net-centric and cloud systems.

The NCSS I/UCRC research will enable industry to develop and leverage emerging net-centric and cloud computing platforms and infrastructures. The industrial sectors that benefit from the center include defense, energy, transportation, health care, commerce, homeland security, and emergency preparedness/emergency response industries. The Center has actively worked with the Metroplex Technology Business Council (MTBC) Innovation Group in Dallas/Fort Worth metropolitan area and led to the creation of a new MTBC Special Interest Group (SIG) in the area of cloud computing. The NCSS I/UCRC will continue to provide education and training of students to meet our nation's future workforce needs in these rapidly emerging and evolving technology areas. Many of the software tools, and infrastructure (e.g., sensors labs) developed at UNT site were made available to students in various courses such as sensor networks (with 30 students per year), computer architecture (with 40 students per year). The UTD site developed courses on cloud computing systems, Internet of Things, programming and optimization of applications on heterogeneous multicore systems, power management for multi-core systems, and other NCSS related topics. Over 60 students have enrolled in these courses over the past 3 years. A significant growth in enrollments in these courses is anticipated. The Center activities also have and will continue to provide real-world research experience for several graduate and undergraduate students as well as international education and research collaboration in this area with some faculty and students from universities in China, France, Germany, Italy, Taiwan, and Turkey.

Satyajeet Nimgaonkar, Mahadevan Gomathisankaran, and Saraju P. Mohanty. "MEM-DnP: A Novel Energy Efficient Approach for Memory Integrity Detection and Protection in Embedded Systems," *Circuits, Systems, and Signal Processing*, v.32, 2013, p. 2581. doi:<http://dx.doi.org/10.1007/s00034-013-9621-4>

Charles Shelor, James Buchanan and Krishna Kavi. "Potential energy savings through eliminating unnecessary writes in the cache-memory hierarchy," *IJCA*, v.21, 2014, p. 178. doi:<http://csrl.cse.unt.edu/~kavi/Research/IJCA-2014>

Chia-en Lin and K. Kavi.. "A QoS-Aware BPEL Framework for Service Selection and Composition Using QoS Properties," *International Journal On Advances in Software*, v.6, 2013, p. 56. doi:<http://csrl.cse.unt.edu/~kavi/Research/IJAS-2013.pdf>

C-Y. Lee, P. Kamongi, K. Kavi and M. Gomathisankaran. "Optimus: Framework of vulnerabilities, attacks, defenses, SLA and Privacy Ontologies," *International Journal of Next-Generation Computing*, v.6, 2015.

K. Kavi, S. Pianelli, G. Pisano and G. Regina.. "Memory organizations for 3D DRAMs and PCMs in processor memory hierarchy," *Journal of Systems Architecture*, v.61, 2015, p. 531. doi:[10.1016/j.sysarc.2015.07.00](https://doi.org/10.1016/j.sysarc.2015.07.00)

M. Rezaei and K. Kavi. "ABT and SBT revisited: Efficient memory management techniques for object-oriented and web-based applications," *International Journal of Science and Technology*, v.23, 2016, p. 1217.

M. Scrbak, M. Islam, K. Kavi, M. Ignatowski, N. Jayasena. "Exploring the processing in memory design space," *Journal of Systems Architecture*, 2016, p. 10.1016/j. doi:[10.1016/j.sysarc.201](https://doi.org/10.1016/j.sysarc.201)

<부록 16> NSF OSS 예비 후보

[NSF 오픈소스소프트웨어 예비후보01]

Chameleon: A Large-Scale, Reconfigurable Experimental Environment for Cloud Research

Collaborative Research:

https://www.chameleoncloud.org/media/cms_page_media/17/Chameleon-fire-geni_OSpyKvd.pdf

Award Number:1743358; Principal Investigator:Katarzyna Keahey; Co-Principal Investigator;; Organization:University of Chicago;NSF Organization:CNS Start Date:10/01/2017; Award Amount:\$3,102,183.00; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1743358

Award Number:1743354; Principal Investigator:Daniel Stanzione; Co-Principal Investigator;; Organization:University of Texas at Austin;NSF Organization:CNS Start Date:10/01/2017; Award Amount:\$2,367,716.00; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1743354&HistoricalAwards=false

Award Number:1743324; Principal Investigator:Joel Mambretti; Co-Principal Investigator;; Organization:Northwestern University;NSF Organization:CNS Start Date:10/01/2017; Award Amount:\$100,000.00; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1743324&HistoricalAwards=false

One of **two** NSFCloud mid-scale experimental computing research infrastructures, **Chameleon** is a **deeply reconfigurable cloud computing testbed** supporting large-scale Computer Science experimentation. Located at the University of Chicago and the Texas Advanced Computing Center, it is composed of over 600 compute nodes and a total of **5 petabytes of storage** supporting data-intensive science. Since its public availability in July 2015, Chameleon has attracted a community of over 1,300 users working on over 200 education and research projects.

In Phase 2 of the NSFCloud program, this project will expand Chameleon capabilities and capacity, and will fulfill its mission of serving as a testbed that is used extensively by the research community by:

- 1) broadening the set of supported experiments through the addition of new hardware and software supporting high bandwidth and Software Defined Networking (SDN)
- 2) providing operational sustainability by streamlining operations, packaging the testbed as open source software that can be leveraged by others, and developing a process for adding resources to the testbed
- 3) provide support for managing, understanding, comparing and reproducing Computer Science experiments, and
- 4) focusing effort on building an engaged community around the testbed via both traditional outreach including targeted training and education efforts, and direct personalized outreach to experimenters.



The Chameleon architecture consists of a set of standard cloud units (SCUs), each of which is a single rack with 42 compute nodes, 4 storage nodes attached to 128TB of local storage in configurable arrays, and an OpenFlow compliant network switch. In addition to the homogeneous SCUs, a variety of heterogeneous hardware types is available to experiment with alternative technologies. The testbed also includes a shared infrastructure with a persistent storage system accessible across the testbed, a top-level network gateway to allow access to public networks, and a set of management and provisioning servers to support user access, control, monitoring and configuration of the testbed. Chameleon is physically distributed between the Texas Advanced Computing Center (TACC) and the University of Chicago (UC) through 100Gbps Internet2 links, to allow users to examine the effects of a distributed cloud.experimenters.

[NSF 오픈소스소프트웨어 예비후보02]

Gunrock: High-Performance GPU Graph Analytics

SI2-SSE:

Award Number:1740333; Principal Investigator:John Owens; Co-Principal Investigator;; Organization:University of California-Davis;NSF Organization:OAC Start Date:10/01/2017; Award Amount:\$400,000.00; Relevance:29.4; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1740333&HistoricalAwards=false

Many sets of data can be represented as "graphs". Graphs express relationships between entities, and those entities and relationships can be used to solve problems of interest in many fields. For instance, a social graph (like Facebook's) links people (entities) by friendships (relationships), and with that graph, Facebook can suggest people to you who might be your friends. Amazon might use a graph made of people and items for sale (entities) connected by who bought those items (relationships) to suggest items you might want to buy. A credit card company might look at your pattern of purchases and detect possible fraud even before you know your credit card was stolen. Graphs are also useful in many fields of science, such as genomics, epidemiology, and economics. This project uses an emerging programmable processor, the graphics processor (GPU), to solve graph problems. GPUs are rapidly moving into our nation's largest data centers and supercomputers. The project team is **building a system for computation on graphs** that will significantly improve performance on these problems. In this project, the team will work with the computing community and the scientific community, both of whom have numerous interesting, challenging graph computation problems that this system will target. The system is open-source software and can be used freely by researchers and industry all over the world.

This project, supported by the Office of Advanced Cyberinfrastructure seeks to develop the **"Gunrock" programmable, high-performance, open-source graph analytics library for graphics processors (GPUs)** from a working prototype to a robust, sustainable, open-source component of the GPU computing ecosystem. Gunrock's strengths are its programming model and highly optimized implementation. With this work the project team hopes to address Gunrock's usability in the computing and scientific communities by improving Gunrock's scalability, capabilities, core operators, and supported graph computations. In this work the team will collaborate with the GPU Open Analytics Initiative and the NSF-sponsored CINET project for network science to ensure that our work has the broadest possible impact.

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.

[NSF 오픈소스소프트웨어 예비후보03]

Entangled Quantum Dynamics in Closed and Open Systems, an Open Source Software Package for Quantum Simulator Development and Exploration of Synthetic Quantum Matter

SI2-SSE:

Award Number:1740130; Principal Investigator:Lincoln Carr; Co-Principal Investigator;; Organization:Colorado School of Mines;NSF Organization:OAC Start Date:09/01/2017; Award Amount:\$499,978.00; Relevance:31.11; August 31, 2020 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1740130&HistoricalAwards=false

On the way to universal quantum computing, **quantum simulators**, literally "**analog**" **quantum computers**, have already been a huge success and are fulfilling Feynman's original 1982 vision of quantum computing. Each such simulator requires a dedicated experimental platform, and costs on the order of several million dollars to build. Such experiments have many interacting parts often requiring a complex rearrangement and months of work in order to perform a specified quantum computation. A widely accessible and easy to use software tool to shortcut design considerations for quantum simulator experimentalists is much needed. This project will **create such a tool**. Initially, the project researchers will design and release on SourceForge an open source software package centered around 1D matrix product state (MPS) and matrix product density operator (MPDO) methods, for both closed and open quantum systems, which any experimentalist can download and easily use locally to design and benchmark their quantum simulator architecture of choice. The software elements will include (i) prebuilt generalized Ising, Hubbard, and other models, (ii) multi-legged ladders to extend into quasi-1D, (iii) different time propagation methods for short and long-range interactions, and (iv) supplemental exact diagonalization and quantum trajectory methods. Secondly, the project members will create an even simpler graphical version via a web interface in collaboration with the Science Gateways Community Institute which allows any experimentalist to run quick simulations and tests on local dedicated high-performance computing resources at the Colorado School of Mines in a secure and user-friendly format. Finally, the project team will develop a key new software element in terms of a series of increasingly accurate discretization schemes to model continuum quantum simulators and mesoscopic limits, as well as control systematic error in experiments.

MPS/MPDO methods enable the treatment of outstanding quantum problems for design of new materials dubbed **synthetic quantum matter**, push the boundaries of quantum mechanics into strongly correlated physics, where particles and even quasiparticles lose meaning, and allow exploration of totally new realms of entangled dynamics. The software package proposed here will make these extraordinary capacities accessible to the over 150 quantum simulator experimental groups, as well as the many theoretical/computational groups investigating entangled quantum dynamics, thus greatly speeding up exploration of quantum simulator physics. The software package will provide a platform for exploring far-from-equilibrium open quantum system dynamical models, an important topic in the theoretical and computational physics community at present due to the new and expanded capacity offered by quantum simulators; and it will provide an educational venue for new graduate students entering research groups and faculty developing courses in computational physics and related areas. Broader impact activities include integration of computation across the Colorado School of Mines undergraduate physics curriculum as a model for departments across the country to achieve this AAPT/APS goal, e.g. redesigning the mathematical physics course to cover equal parts analytical and computational methods, including specific research skills for summer REUs and internships. Finally, graduate students will be trained in critical analysis for scientific problem solving and rigorous numerical techniques for high-performance computing, including open source science via a science gateway, an approach key to success in a number of arenas in society, from the materials genome initiative to the space program.

This project is supported by the Office of Advanced Cyberinfrastructure in the Directorate for Computer & Information Science and Engineering and the Physics Division, Materials Research Division, and Office of Multidisciplinary Activities in the Directorate of Mathematical and Physical Sciences.

[NSF 오픈소스소프트웨어 예비후보04]

An Ecosystem of Reusable Image Analytics Pipelines

SI2-SSE:

Award Number:1739419; Principal Investigator:Andrew Connolly; Co-Principal Investigator:Mario Juric, Ariel Rokem, Alvin Cheung, Magdalena Balazinska; Organization:University of Washington;NSF Organization:OAC Start Date:09/01/2017; Award Amount:\$500,000.00; Relevance:29.06; August 31, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1739419&HistoricalAwards=false

Astronomy has entered an era of massive data streams generated by telescopes and surveys that can scan tens of thousands of square degrees of the sky across many decades of the electromagnetic spectrum. The promise of these new experiments - characterizing the nature of dark energy and the composition of dark matter, discovering the most energetic events in the universe, tracking asteroids whose orbits may intersect with that of the Earth - will only be realized if we can address the challenge of how to process and analyze the tens of petabytes of images that these astronomical surveys will generate per year. With the increasing capacity for scientists to collect ever larger sets of data, often in the form of images, our potential for scientific discovery will soon be limited not by how we collect or store data, but rather how we extract the knowledge that these data contain (e.g. how we account for noise inherent within the data, and understand when we have detected fundamentally new classes and interesting events or physical phenomena). This project is **to develop an open source scalable framework for the analysis of large imaging data sets**. It is designed to operate as a cloud service, incorporate seamlessly new or legacy image processing algorithms, support and optimize complex analysis workflows, and scale analyses to thousands of processors without the need for an individual user to develop custom solutions for a specific computer platforms or architecture. This framework will be integrated with state-of-the-art image analysis algorithms developed for astronomical surveys to provide an image analytics platform that can be used by future telescopes and cameras and the astronomical community as a whole. **Beyond astronomy**, the framework will be extended to enable scientists from the physical and life sciences that make use of imaging data (e.g. neuroscience, oceanography, biology, seismology) to focus their work on developing scientific algorithms and analyses rather than the infrastructure required to process massive data sets

Over the last decade, there have been many advancements in astronomical image analysis algorithms and techniques; driven by new surveys and experiments. The complexity of these techniques and the systems that run them has, however, meant that the number of users who make use of these advancements is small; typically restricted to the experiments themselves or to a small group of expert users. Because of this, the community as a whole does not benefit from the significant investment in image analytics for astronomy. In this project, the PIs address these issues by developing and deploying a scalable framework for the analysis of small and large imaging datasets. This cloud-based system will be able to incorporate new and legacy image processing algorithms, support and optimize complex analysis workflows, scale applications to thousands of processors without users needing to develop custom code for specific platforms, and support efficient sharing of algorithms and analysis results among users. It will enable state-of-the-art image analysis algorithms (e.g. those developed for surveys such as the Large Synoptic Survey Telescope; LSST) to be used by the broad astronomical community and in so doing will leverage then tens of thousands of hours that has been invested in the development of these techniques. To accomplish this the team will extract key data analysis functions from the LSST data analysis pipeline into a standalone library, independent of the LSST software stack and data access mechanisms. They will integrate this library with the **Myria** big data management system. Myria is an elastically scalable big data management system that operates as a service in the Amazon cloud that we developed at the University of Washington. Compared with other big data systems, Myria is especially attractive because it integrates PostgreSQL database instances within its storage layer and thus provides access to PostgreSQL's rich libraries of spatial functions, which are frequently used in astronomical data analysis pipelines. At the same time, it has rich support for new and legacy Python code and for complex analytics. By integrating the library of LSST image analytics functions with Myria, new image analytics pipelines will become significantly easier to write. The skeleton of the analysis pipeline will be expressed in the MyriaL declarative query language (i.e. SQL extended with constructs such as iterations and others). The core data processing functions will directly map to Python functions, enabling the reuse of legacy code and the easy addition of new functions. The resulting code will be amenable to optimization and efficient execution using the Myria service. By doing so, they intend to reduce barriers to adoption. Users will be able to express their analysis in Python without worrying about how data and computation will be distributed in a cluster. The image analysis framework developed as part of this proposal will be made publicly available as **open-source software**. The PIs will utilize the use case of neuroscience to demonstrate how their system, developed for astronomy, can be deployed across multiple domains.

This project is supported by the Office of Advanced Cyberinfrastructure in the Directorate for Computer & Information Science and Engineering, the Astronomical Sciences Division and Office of Multidisciplinary Activities in the Directorate of Mathematical and Physical Sciences.

[NSF 오픈소스소프트웨어 예비후보05]

Enhancing Context-Awareness and Personalization for Intensively Adaptive Smoking Cessation Messaging Interventions

SCH: INT: Collaborative Research:

Award Number:1722792; Principal Investigator:Benjamin Marlin; Co-Principal Investigator:Prashant Shenoy, Deepak Ganesan; Organization:University of Massachusetts Amherst;NSF Organization:IIS Start Date:10/01/2017; Award Amount:\$640,195.00; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1722792&HistoricalAwards=false

Award Number:1722646; Principal Investigator:Santosh Kumar; Co-Principal Investigator;; Organization:University of Memphis;NSF Organization:IIS Start Date:10/01/2017; Award Amount:\$196,446.00; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1722646&HistoricalAwards=false

Award Number:1722708; Principal Investigator:Rajani Shankar Sadasivam; Co-Principal Investigator:Thomas Houston; Organization:University of Massachusetts Medical School;NSF Organization:IIS Start Date:10/01/2017; Award Amount:\$163,356.00; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1722708&HistoricalAwards=false

According to the Centers for Disease Control, tobacco use remains the leading preventable cause of death in the US, causing approximately 480,000 deaths each year, and incurring over \$150 billion in healthcare costs. As a result, scalable, low cost, and effective smoking cessation interventions are clearly needed. This research aims **to develop and validate new messaging-based smoking cessation support intervention systems** that will leverage recent advances in smart wearable technologies to significantly enhance efficacy. The system will integrate wearable sensors that continuously estimate an individual's level of stress and craving as well as the occurrence of smoking. This information will be used to enhance the context awareness of the intervention system, allowing it to continuously adapt both the content and delivery timing of intervention components for each individual. By developing scalable messaging-based smoking cessation support interventions with improved personal relevance, this research has the potential to lead to direct benefits to society by more effectively helping individuals to quit smoking.

To accomplish the goal of providing effective, personalized smoking cessation interventions, this research will develop and evaluate the models, algorithms, and wearable-phone-cloud computational infrastructures required to support the context inferences, personalization, and delivery timing optimizations required. Starting from the team's extensive prior work, this research will contribute to (1) advances in mobile health sensing and context inference with low-cost, low-power sensors; (2) advances in real-time, stream-based active learning for personalizing context inference models; (3) advances in contextualized recommender systems to personalize message selection based on inferred contexts; and (4) advances in robust, real-time wearable-phone-cloud data analytics systems. This work will also make substantial contributions to enhancing research infrastructure through **open source software** releases that can be leveraged by the research community to yield benefits in other high-profile health areas including heart disease, obesity, and addiction.

[NSF 오픈소스소프트웨어 예비후보06]

Flexible Resource Management and Coordination Schemes for Lightweight, Rapidly Deployable OS/Rs

CSR: Small: Collaborative Research:

Award Number:1718252; Principal Investigator:Kyle Hale; Co-Principal Investigator;; Organization:Illinois Institute of Technology;NSF Organization:CNS Start Date:08/01/2017; Award Amount:\$249,771.00; Relevance:29.28; July 31, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1718252&HistoricalAwards=false

Award Number:1718287; Principal Investigator:John Lange; Co-Principal Investigator;; Organization:University of Pittsburgh;NSF Organization:CNS Start Date:08/01/2017; Award Amount:\$249,964.00; Relevance:29.47; July 31, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1718287&HistoricalAwards=false

Current cloud systems leverage either heavy-weight virtualization (running applications inside full-fledged virtual machines (VMs) with their own operating systems) or containers (light-weight software environments that share a single underlying operating system). This collaborative project aims to increase the efficiency of cloud systems by **enabling the rapid deployment of specialized operating system and runtime (SOS/R) environments optimized for particular applications or runtime systems**. This will allow users of applications with special performance requirements to enjoy the convenience of cloud infrastructure without sacrificing performance or design constraints imposed by general-purpose operating systems.

The project will advance the state of the art in cloud system software by: (a) enabling the development of a general specialized hosting framework allowing the deployment of arbitrary system software stacks as both VMs and native co-kernel instances; (b) a set of interfaces and mechanisms to allow SOS/R environments to react to highly dynamic and variable hardware resource availability; (c) novel approaches that leverage hardware virtualization features to host SOS/Rs; and (d) a thorough evaluation of all of these approaches using realistic cloud workloads at scale.

This work will enable users of cloud systems to efficiently deploy diverse and specialized system architectures. This will not only enhance the capabilities of existing cloud and data center platforms, but will also enable the deployment of new types of operating systems and runtimes. Prototype systems produced in this effort will be made available to the general public as open-source software, and will also be leveraged as teaching and platforms for courses in operating systems, virtualization, runtime environments, and general cloud computing.

Data, code, and results generated from the project will be made freely available indefinitely on the team's webpages. All software packages will be accompanied by instructions for building and running them using freely available open-source tools. Data and results will be made available in compressed archive formats. Links to specific repositories (such as git repositories for code) will be made public either at <http://hexsa.halek.co> or <http://prognosticlab.org>.

[NSF 오픈소스소프트웨어 예비후보07]

Creating Natural Data Visualization and Analysis Environments

III: Small:

Award Number:1717111; Principal Investigator:John Stasko; Co-Principal Investigator;; Organization:Georgia Tech Research Corporation;NSF Organization:IIS Start Date:11/15/2017; Award Amount:\$493,752.00; Relevance:28.3; October 31, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1717111&HistoricalAwards=false

In the current data-rich era, data visualization and exploration capabilities are becoming more widely used in a variety of disciplines including business, health, education, and public policy, among others. Currently, people use visualization systems on desktop and laptop computers and typically interact via keyboard and mouse. Such interactions, while useful, pale in comparison to the natural, fluid interactions presented in futuristic feature films such as "**Minority Report**" and "**Iron Man**" where characters interact with large, projected wall displays through speech, gaze, and gesture. To move towards such futuristic interfaces, we must develop **new forms of Natural User Interfaces (NUIs)** employing **multimodal interactions** such as speech, pen, touch, gestures, gaze, and head and body movements. While no specific interaction modality may provide all desired capabilities, combinations of modalities (e.g., speech, gaze, and pen) could ideally provide a more natural, intuitive, and integrated interface experience. This project will **explore, design, develop, and evaluate NUIs for data visualization and visual analytics**. Developing techniques and systems that provide natural, expressive, multimodal input and interaction for multiple representations of data has the potential to broadly impact a virtually unlimited number of disciplines and areas of society. Additionally, the project will provide educational experience and research training for graduate and undergraduate students to help assist them in their careers.

Drawing upon prior research on natural language and multi-touch interfaces for visualization, the research seeks to enable a next generation of powerful, expressive, and natural systems that facilitate fluid interaction with visual representations of data. While these interaction modalities harbor great potential, many research challenges exist and must be addressed. For instance, how should a system handle speech input that is ill-formed, incomplete, or ambiguous? What if an intention is misinterpreted or misunderstood? Regarding multi-touch gesture input, what are the "best" touch gestures to use in these interfaces? Do those gestures map well to different types of visualizations and to different display types and sizes and types? How do we make these gestures easier to discover and learn? Combining multiple input modalities may allow systems to counterbalance the weaknesses of one modality with the strengths of another (e.g., ambiguity in selection via speech can be balanced by the preciseness of touch), facilitating a naturalistic user experience. Project objectives include the design, implementation, and evaluation of multimodal interfaces to data visualization systems. In particular, the research will investigate how different interaction methods affect and enable data analysis and exploration. The project will create applications for particular data domains as well as open-source toolkits for other researchers to use in their own work. User studies will identify if and how different types of interaction (speech, touch, gesture, etc.) make analysis faster, easier to learn, and more powerful. The project website (<http://www.cc.gatech.edu/gvu/ii/naturalvis>) will include project information, links to resulting publications, videos, and open-source software that is produced.

[NSF 오픈소스소프트웨어 예비후보08]

Enabling Sensor-Rich Vehicular Applications with Edge Computing

CSR:Small:

Award Number:1717064; Principal Investigator:Jason Flinn; Co-Principal Investigator;; Organization:University of Michigan Ann Arbor;NSF Organization:CNS Start Date:10/01/2017; Award Amount:\$499,387.00; Relevance:28.12; September 30, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1717064&HistoricalAwards=false

This project is **developing methodologies and systems that support sensor-rich vehicular applications by offloading computation to edge compute resources located at roadside and embedded within cellular networks**. The modern vehicle has cameras, microphones, location and positioning sensors that all supplement detailed monitoring of vehicle motion and operation. These sensors can supply data for advanced applications, such as personal driving assistants, automated road hazard detection, and augmented display of driving data (e.g., highlighting turn-by-turn directions or available parking on the vehicle windshield). However, sensor-rich applications also require substantial computing power that is unavailable in the vehicle. These applications cannot leverage cloud resources because the latency of communicating with remote data centers is too high and the bandwidth available over wide-area networks is too limited. Instead, this project is supporting sensor-rich applications by **offloading their computation to nearby servers located at the network edge**. This work has the potential to enable applications with great societal benefit. Sensor-rich vehicular applications promise to improve driving efficiency; for example, by reducing congestion and idling time by guiding drivers directly to available parking spots. These applications can also substantially improve road safety; for example, by providing automated, crowd-sourced road hazard detection. The project is prototyping these applications and developing the infrastructure required for their deployment. The research results, traces of vehicular applications and environments, and open-source software artifacts developed by this project are available at:

<http://pervasive.eecs.umich.edu/vehicular-edge.html>.

Since vehicles will remain near edge resources for only limited periods of time, the project is creating methods for fast handover of state and computation from one edge resource to another. Vehicle mobility introduces considerable variability in network and service performance, so the project is investigating how the principled use of redundancy can improve average and worst-case response time. Finally, to support applications with extremely tight time bounds, the project is using speculative execution to hypothesize multiple possible futures that depend on vehicle speed and direction, calculate results based on those futures, and then select the appropriate response based upon actual vehicle motion during the time when calculations were made.

[NSF 오픈소스소프트웨어 예비후보09]

The SimCardio Open Source Multi-Physics Cardiac Modeling Package

SI2-SSI Collaborative Research:

Award Number:1663671; Principal Investigator:Alison Marsden; Co-Principal Investigator;;
Organization:Stanford University;NSF Organization:OAC Start Date:09/01/2017; Award
Amount:\$1,431,169.00; Relevance:31.08; August 31, 2021 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1663671&HistoricalAwards=false

Award Number:1663747; Principal Investigator:Shawn Shadden; Co-Principal Investigator;;
Organization:University of California-Berkeley;NSF Organization:OAC Start Date:09/01/2017; Award
Amount:\$943,832.00; Relevance:30.69; August 31, 2021 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1663747&HistoricalAwards=false

Cardiovascular (CV) simulations have become a crucial component of fundamental research in surgical planning, device design, diagnosis, and disease mechanisms. The project team has previously developed **SimVascular** (www.simvascular.org), which is currently the only open source software package providing a complete pipeline from medical image data segmentation to patient specific blood flow simulation and analysis in arteries and veins. The **SimCardio** open source project will extend and enhance the functionality of SimVascular to the realm of heart modeling, providing the first fully integrated computer model of cardiac physiology and function. This will help basic science and medical researchers perform computer modeling in numerous diseases affecting heart function in children and adults. This computer modeling software will enable researchers to build models of the heart and vascular anatomy directly from medical imaging data, which can be used for personalized treatment planning and medical device design, ultimately leading to new treatments for patients with cardiovascular disease.

The SimCardio project will create a unique open source software package for multi-physics cardiac modeling and simulations. SimCardio will include a new multi-physics finite element solver with capabilities for large-deformation fluid-structure interaction (FSI) to capture ventricular contraction and heart valve dynamics, non-linear and visco-elastic material models, cardiac mechanics models of active heart contraction, and electrophysiology. This will be facilitated by sustainable software infrastructure bridging the cardiovascular fluid and solid mechanics communities. The project will provide a new user-interface for high-throughput construction of patient-specific cardiac and vascular models. SimCardio will broaden the applicability of SimVascular to problems including heart valves, heart failure, cardiomyopathy, aortic dissection, structural congenital heart defects and medical devices. To facilitate adoption, the project will publicly provide accompanying educational and training materials, and maintain a sustainable software ecosystem to increase the user community and ensure continued availability and evolution.

[NSF 오픈소스소프트웨어 예비후보10]

BONSAI: An Open Software Infrastructure for Parallel Autotuning of Computational Kernels

SI2-SSE:

Award Number:1642441; Principal Investigator:Jakub Kurzak; Co-Principal Investigator:Piotr Luszczek; Organization:University of Tennessee Knoxville;NSF Organization:OAC Start Date:11/01/2016; Award Amount:\$499,977.00; Relevance:30.55; October 31, 2019 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1642441&HistoricalAwards=false

Most supercomputers today accelerate the computations by using processors with many cores to solve important problems in science and engineering. Although this reduces the cost of the hardware system, it greatly increases the complexity of writing and optimizing ("tuning") software. This project extends a previously funded NSF project: Benchmarking Environment for Automated Software Tuning (BEAST) program to create a software toolkit that allows for semi-automatic optimization of software, thereby reducing the programming overhead. This project, **BEAST Open Software Autotuning Infrastructure (BONSAI)** will greatly increase the efficiency of scientists and engineers to develop fast and efficient programs to solve their problems. BONSAI has tremendous support from various computer processor manufacturing companies and academic institutions. The BONSAI system will be available as open-source software for academic and commercial use and many students will be trained in using the software.

The emergence and growing dominance of hybrid systems that incorporate accelerator processors, such as GPUs and coprocessors, have made it far more difficult to optimize the performance of the different computational kernels that do the majority of the work in most research applications. The BONSAI project aims to create a transformative solution to this problem by developing a software infrastructure that uses parallel hybrid machines to enable large autotuning sweeps on computational kernels for GPU accelerators and many-core coprocessors. The system will go beyond just measuring runtimes, allowing for collection and analysis of non-trivial amount of data from hardware performance counters and power meters. The system will have a modular architecture and rely on standard data formats and interfaces to easily connect with mainstream tools for data analytics and visualization. The BONSAI project will leverage the experiences of the BEAST project, which established a successful autotuning methodology and validated an autotuning workflow. BONSAI will equip the community with a software environment for applying parallel resources to the tuning and performance analysis of computational kernels. Specifically, the work will be organized around the following objectives: (1) Harden and extend the programming language called BeastLang, which was created during prior research as a way of defining the search space that the autotuning infrastructure generates and explores. BeastLang enables users to create parameterized kernel specifications that encode the interplay between the kernels themselves, the compilation tools, and the target hardware. It will be integrated with the other components of BONSAI, have its Python syntax enhanced and extended, its compiler improved, and be supplemented with a runtime that supports it with multi-way parallelism for the autotuning process. (2) Develop and test a benchmarking engine for making large scale parallel tuning sweeps, using large numbers of GPU accelerators or many-core coprocessors. This engine will support both parallel compilation and parallel tests of the resulting kernels, using many distributed memory nodes and multithreading within each node, with dynamic load balancing. It will produce an extensive collection of performance information from hardware counters, and possibly energy meters, as well as collection of information about the saturation of the compiled code with different classes of instructions. (3) Develop and test a software infrastructure for collecting, preprocessing, and analyzing BONSAI performance data. The system will a) simplify the task of instrumenting the kernel and provide a simple interface for selecting the metrics to be collected with sensible defaults; b) simplify the process of collecting hardware counters and performance data from various open source and vendor specific libraries; and c) provide tools that allow the user to quickly and efficiently transform output data to a format that can be easily read and analyzed using mainstream tools such as R and Python. (4) Document and illustrate the process of using BONSAI to tune various different types of kernels. These model case studies will include discussions of how BeastLang was applied to create the parameterized kernel stencil, how the parallel benchmarking engine is invoked to generate and explore the search space, and how the data collected from the operation of the engine can be analyzed and visualized to gain insights that can correct or refine the process for another iteration. The BONSAI project has the potential to fundamentally transform autotuning research by: 1) Making autotuning accessible to a broad audience of developers from a broad range of computing disciplines, as opposed to a few selected individuals with the wizardry to set up a successful experiment within the confines of serial execution. 2) Changing the general perception of autotuning as not just the means of producing fast code, but as a general technique for performance analysis and reasoning about the complex software and hardware interactions, and positioning the technique as one of primary tools for hardware-software co-design. 3) Boosting interest in exploring neglected avenues of computing, such as exploration of unorthodox data layouts, and challenge the status quo of legacy software interfaces. BONSAI has the potential to bring autotuning to the forefront of software development and to help position autotuning as a pillar of software engineering

[NSF 오픈소스소프트웨어 예비후보11]

Software Framework for Research in Semi-Autonomous Teleoperation

NRI: Collaborative Research:

Award Number:1637789; Principal Investigator:Peter Kazanzides; Co-Principal Investigator:Russell Taylor; Organization:Johns Hopkins University;NSF Organization:IIS Start Date:10/01/2016; Award Amount:\$978,300.00; Relevance:29.49; September 30, 2019 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1637789&HistoricalAwards=false

Award Number:1637759; Principal Investigator:Gregory Fischer; Co-Principal Investigator;; Organization:Worcester Polytechnic Institute;NSF Organization:IIS Start Date:10/01/2016; Award Amount:\$550,157.00; Relevance:29.39; September 30, 2019 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1637759&HistoricalAwards=false

Award Number:1637444; Principal Investigator:Blake Hannaford; Co-Principal Investigator;; Organization:University of Washington;NSF Organization:IIS Start Date:10/01/2016; Award Amount:\$479,998.00; Relevance:29.49; September 30, 2019 (Estimated)
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1637444&HistoricalAwards=false

Telemanipulation systems consist of a human interacting with a mechanical device on the master side to operate a robot at the remote side. They provide natural opportunities for research in intelligent human/robot collaboration, but existing commercial systems, used in areas such as telesurgery, are not intelligent and therefore only replicate the actions of the human operator. These systems are also proprietary, expensive, and not available for modification by researchers. The goal of this NRI project is to provide an **open-source software infrastructure** that is designed to work with **a broad range of hardware and simulated devices** to enable a larger community to pursue research and education in intelligent telemanipulation **at a lower cost**.

The increasing pace of robotics research can be attributed, at least in part, to the increasing availability of software infrastructure, such as **Robot Operating System (ROS)**, and **open hardware platforms**. This NRI project focuses on providing a software infrastructure for research in intelligent telemanipulation, leveraging infrastructure developed for the **Raven II robot** and the **da Vinci Research Kit (dVRK)** and continuing to extend it to **other systems**, including simulated robots. The three main tasks are to: (1) engage the community to guide development, (2) develop and implement a common API for the diverse hardware platforms, and (3) provide a set of high-level, platform-independent software modules. The goal is to support research towards semi-autonomous telerobotic systems that can more effectively combine the knowledge, reasoning, and decision-making capabilities of a human with the sensing and manipulation capabilities of a robot.

[NSF 오픈소스소프트웨어 예비후보12]

Robust Algorithms for an Open Source Software Reliability Tool

CSR: SMALL:

Award Number:1526128; Principal Investigator:Lance Fiondella; Co-Principal Investigator;;
Organization:University of Massachusetts, Dartmouth;NSF Organization:CNS Start Date:10/01/2015;
Award Amount:\$137,554.00; Relevance:31.18; September 30, 2018 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1526128&HistoricalAwards=false

The key to the success of all software is its reliability. This project will develop an **open source software reliability tool** that will allow software engineers to automatically apply software reliability models to help organizations ensure that software applications they develop can operate free of failures. Traditional algorithms are numerically unstable, meaning that they can fail if initial estimates are inaccurate. Failure of an algorithm renders it impossible to apply a software reliability model to make useful predictions such as the amount of additional time a software application should be tested in order to achieve a desired level of reliability.

The goal of this research is to develop **numerically stable algorithms** that will succeed even if the initial estimates are inaccurate. Expectation maximization (EM) and expectation conditional maximization (ECM) algorithms will be developed for failure rate and nonhomogeneous Poisson process (NHPP) software reliability models. Traditional EM algorithms impose restrictive assumptions that limit their application to only the simplest models, while the potential of the ECM algorithm has not been fully explored. Therefore, this research challenge will remove the restrictions of existing EM algorithms and design efficient ECM algorithms for software reliability models. Implementations of these numerically stable EM and ECM algorithms will be incorporated into the open source tool to ensure that software reliability models can be applied successfully. The enhanced stability of the algorithms and the open source nature of the tool may promote widespread use of quantitative software reliability models, enabling companies and organizations to improve time to market or field a software product.

[NSF 오픈소스소프트웨어 예비후보13]

Enabling Technologies for 21st Century Entity Matching Applications

III: Medium:

Award Number:1564282; Principal Investigator:AnHai Doan; Co-Principal Investigator:Paul Hanson, Miron Livny, Jeffrey Naughton; Organization:University of Wisconsin-Madison;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$1,085,906.00; Relevance:28.3; August 31, 2020 (Estimated)

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1564282&HistoricalAwards=false

Entity matching (EM) decides if different data instances (such as "**UW-Madison**" and "**Univ of Wisc Madison**") refer to the same real-world entity. This problem is critical for numerous Big Data and data science applications. This project will develop solutions and tools for EM which will significantly advance the state of the art. Compared to the current solutions, the proposed solutions will consider the entire EM pipeline, will be usable by lay users (such as domain scientists and journalists), will scale to large amounts of data, and will exploit crowdsourcing to maximize matching accuracy. Technologies developed in this project will be evaluated in three domains: managing scientific data for limnology (the study of lakes and other bodies of freshwater), product matching for e-commerce at WalmartLabs, and developing the Internet of Buildings at Johnson Control Inc. As such, the project will facilitate the widespread deployment of EM tools, thus resulting in more effective information management and access for society. Through its release of **open-source software** it will help educate next-generation workers and researchers. The research will help domain scientists in limnology, and can potentially impact hundreds of thousands of buildings and millions of users via collaboration with Johnson Controls and WalmartLabs. Finally, a planned textbook and open-source system artifacts from the project will be disseminated broadly in the research community, to significantly enhance the data management infrastructure for research and education.

The project will introduce both conceptual and technical novelties. Conceptually, instead of focusing on just the matching step (and studying how to match accurately and scalably, as much of current work has done), the project advocates developing solutions for the entire raw-data-to-matches EM pipeline. Further, it advocates using the matching step to guide the execution of the remaining steps in the EM pipeline. Technically, the project will develop novel solutions for non-matching steps in the EM pipeline, and do so in a matching-driven fashion. It also introduces new important problems, such as EM debugging. Finally, it develops novel solutions to scale up the entire EM pipeline and to exploit crowdsourcing. As described, the project takes the next logical step in EM research, and can significantly advance the state of the art. Further, many problems underlying this research have commonalities with other data management scenarios. Hence, the research has the potential to contribute to those areas as well. For more information, see the project's homepage at <https://sites.google.com/site/anhaidgroup/projects/nsf-em-project-2016>.

[NSF 오픈소스소프트웨어 예비후보14]

Open Gateway Computing Environments Science Gateways Platform as a Service (OGCE SciGaP)

Collaborative Research: SI2-SSI:

Award Number:1339774; Principal Investigator:Marlon Pierce; Co-Principal Investigator:Suresh Marru; Organization:Indiana University;NSF Organization:OAC Start Date:10/01/2013; Award Amount:\$**2,519,986.00**; Relevance:30.61;
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1339774&HistoricalAwards=false

Award Number:1339856; Principal Investigator:Mark Miller; Co-Principal Investigator:Amitava Majumdar; Organization:University of California-San Diego;NSF Organization:OAC Start Date:10/01/2013; Award Amount:\$**1,742,099.00**; Relevance:30.61;
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1339856&HistoricalAwards=false

Award Number:1339649; Principal Investigator:Barries Demeler; Co-Principal Investigator;; Organization:University of Texas Health Science Center San Antonio;NSF Organization:OAC Start Date:10/01/2013; Award Amount:\$**600,065.00**; Relevance:30.81;
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1339649&HistoricalAwards=false

Science Gateways are **virtual environments** that dramatically accelerate scientific discovery by enabling scientific communities to utilize distributed computational and data resources (that is, cyberinfrastructure). Successful Science Gateways provide access to sophisticated and powerful resources, while shielding their users from the resources' complexities. Given Science Gateways' demonstrated impact on progress in many scientific fields, it is important to remove barriers to the creation of new gateways and make it easier to sustain them. The **Science Gateway Platform (SciGaP) project** will create **a set of hosted infrastructure services** that can be easily adopted by gateway providers to build new gateways based on robust and reliable open source tools. The proposed work will transform the way Science Gateways are constructed by significantly lowering the development overhead for communities requiring access to cyberinfrastructure, and support the efficient utilization of shared resources.

SciGaP will **transform access to large scale computing and data resources** by reducing development time of new gateways and by accelerating scientific research for communities in need of access to large-scale resources. SciGaP's adherence to open community and open governance principles of the Apache Software Foundation will assure open source software access and open operation of its services. This will give all project stakeholders a voice in the software and will clear the proprietary fog that surrounds cyberinfrastructure services. The benefits of SciGaP services are not restricted to scientific fields, but can be used to accelerate progress in any field of endeavor that is limited by access to computational resources. SciGaP services will be usable by a community of any size, whether it is an individual, a lab group, a department, an institution, or an international community. SciGaP will help train a new generation of cyberinfrastructure developers in open source development, providing these early career developers with the ability to make publicly documented contributions to gateway software and to bridge the gap between academic and non-academic development.

<부록 17> H2020 CPS 예비 후보

[H2020 CPS 예비후보01]

Dependability Engineering Innovation for CPS (DEIS)

http://cordis.europa.eu/project/rcn/205982_en.html

Project ID: 732242

From **2017-01-01 to 2019-12-31**, ongoing project

Total cost: EUR 4 889 290 (EU contribution: EUR 4 889 290)

Coordinated in: Austria AVL LIST GMBH

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-01-2016 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2016-1

Funding scheme: RIA - Research and Innovation action

Objective

Cyber-Physical-Systems harbor the potential for vast economic and societal impact in all major application domains, however in case of failure this may lead to catastrophic results for industry and society. Thus, ensuring the dependability of such systems is the key to unlocking their full potential and enabling European industries to develop confidently business models that will nurture their societal uptake.

The DEIS project addresses this challenges by developing technologies that form a science of **dependable system integration**. In the core of these technologies lies the concept of a **Digital Dependability Identity (DDI)** of a component or system. DDIs are composable and executable in the field facilitating (a) efficient synthesis of component and system dependability information over the supply chain and (b) effective evaluation of this information in-the-field for safe and secure composition of highly distributed and autonomous CPS. This concept shall be deployed and evaluated in four use cases:

Automotive: Stand-alone system for intelligent physiological parameter monitoring

Automotive: Advanced driver simulator for evaluation of automated driving functions

Railway: Plug-and-play environment for heterogeneous railway systems

Healthcare: Clinical decision support app for oncology professionals

The DEIS project will impact the CPS market by providing new engineering methods and tools reducing significantly development time and cost of ownership, while supporting integration and interoperability of dependability information over the product lifecycle and over the supply chain. The development and application of the DDI approach on four use cases from three different application domains will illustrate the applicability of the DDI concept while increasing the competitiveness of the use case owners in their respective markets.

[H2020 CPS 예비후보02]

Platform for Open Development of Systems of Artificial Intelligence (Bonseyes)

http://cordis.europa.eu/project/rcn/206356_en.html

Project ID: 732204

From 2016-12-01 to 2019-11-30, ongoing project

Total cost: EUR 8 593 952,50 (EU contribution: EUR 5 018 025)

Coordinated in: Switzerland NVISO SA

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-01-2016 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2016-1 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

The Bonseyes project aims to develop **a platform** consisting of a **Data Marketplace**, **Deep Learning Toolbox**, and **Developer Reference Platforms** for organizations wanting to adopt Artificial Intelligence in **low power IoT devices** ("edge computing"), **embedded computing systems**, or **data center servers** ("cloud computing"). It will bring about orders of magnitude improvements in efficiency, performance, reliability, security, and productivity in the design and programming of Systems of Artificial Intelligence that incorporate Smart Cyber Physical Systems while solving a chicken-egg problem for organizations who lack access to Data and Models. It's open software architecture will facilitate adoption of the whole concept on a wider scale.

It aims to address one of the most significant trends in the Internet of Things which is the shifting balance between edge computing and cloud computing. The early days of the IoT have been characterized by the critical role of cloud platforms as application enablers. Intelligent systems have largely relied on the cloud level for their intelligence, and the actual devices of which they consist have been relatively unsophisticated. This old premise is currently being shaken up, as the computing capabilities on the edge level advance faster than those of the cloud level. This paradigm shift—from the connected device paradigm to the intelligent device paradigm opens up numerous opportunities.

To evaluate the effectiveness, technical feasibility, and to quantify the real-world improvements in efficiency, security, performance, effort and cost of adding AI to products and services using the Bonseyes platform, four complementary demonstrators will be built: Automotive Intelligent Safety, Automotive Cognitive Computing, Consumer Emotional Virtual Agent, and Healthcare Patient Monitoring. Bonseyes platform capabilities are aimed at being aligned with the **European FI-PPP** activities and take advantage of its flagship project **FIWARE**.

[H2020 CPS 예비후보03]

Unifying Control and Verification of Cyber-Physical Systems (UnCoVerCPS)

http://cordis.europa.eu/project/rcn/194114_en.html

http://cordis.europa.eu/result/rcn/190443_en.html

Project ID: 643921

From 2015-01-01 to 2018-12-31, ongoing project

Total cost: EUR 4 932 902,75 (EU contribution: EUR 4 932 902,25)

Coordinated in: Germany TECHNISCHE UNIVERSITAET MUENCHEN

Funded under: H2020-EU.2.1.1.1. - A new generation of components and systems:

Engineering of advanced embedded and energy and resource efficient components and systems

Topic(s): ICT-01-2014 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2014-1 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

The proposed research effort provides methods for a faster and more efficient development process of **safety- or operation-critical cyber-physical systems** in (partially) unknown environments. Cyber-physical systems are very hard to control and verify because of the mix of discrete dynamics (originating from computing elements) and continuous dynamics (originating from physical elements). We present completely new methods for de-verticalisation of the development processes by a generic and holistic approach towards reliable cyber-physical systems development with formal guarantees. In order to guarantee that specifications are met in unknown environments and in unanticipated situations, we synthesise and verify controllers on-the-fly during system execution. This requires to unify control and verification approaches, which were previously considered separately by developers. For instance, each action of an automated car (e.g. lane change) is verified before execution, guaranteeing safety of the passengers. We will develop completely new methods, which are integrated in tools for modelling, control design, verification, and code generation that will leverage the development towards reliable and at the same time open cyber-physical systems. Our approach leverages future certification needs of open and critical cyber-physical systems. The impact of this project is far-reaching and long-term: UnCoVerCPS prepares the EU to be able to develop critical cyber-physical systems that can only be realised and certified when uncertainties in the environment are adequately considered. This is demonstrated by applying our ground-breaking methods to automated vehicles, human-robot collaborative manufacturing, and smart grids within a consortium that has a balanced participation of academic and industrial partners.

[H2020 CPS 예비후보04]

SAFeTy and secURity by design for interconnected mixed-critical cyber-physical systems (SAFURE)

http://cordis.europa.eu/project/rcn/194149_en.html

http://cordis.europa.eu/result/rcn/192897_en.html

Project ID: 644080

From 2015-02-01 to 2018-01-31, ongoing project

Total cost: EUR 5 702 631,25 (EU contribution: EUR 5 231 375)

Coordinated in: Austria

TECHNIKON FORSCHUNGS- UND PLANUNGSGESELLSCHAFT MBH

Funded under: H2020-EU.2.1.1.1. - A new generation of components and systems:

Engineering of advanced embedded and energy and resource efficient components and systems

Topic(s): ICT-01-2014 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2014-1See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

SAFURE targets the design of cyber-physical systems by implementing a methodology that ensures **safety** and **security** "by construction". This methodology is enabled by a framework developed to extend system capabilities so as to control the concurrent effects of security threats on the system behaviour.

The current approach for security on safety-critical embedded systems is generally to keep subsystems separated, but this approach is now being challenged by technological evolution towards **openness**, **increased communications** and **use of multi-core architectures**.

The objectives of SAFURE are to (1) implement a holistic approach to safety and security of embedded dependable systems, preventing and detecting potential attacks; (2) to empower designers and developers with analysis methods, development tools and execution capabilities that jointly consider security and safety; (3) to set the ground for the development of SAFURE-compliant mixed-critical embedded products.

The results of SAFURE will be (1) a framework with the capability to detect, prevent and protect from security threats on safety, able to monitor from application level down to the hardware level potential attacks to system integrity from time, energy, temperature and data threats; (2) a methodology that supports the joint design of safety and security of embedded systems, assisting the designer and developers with tools and modelling languages extensions; (3) proof-of concept through 3 industrial use cases in automotive and telecommunications; (4) recommendations for extensions of standards to integrate security on safety-critical systems; (5) specifications to design and develop SAFURE-compliant products.

The impact of SAFURE will help European suppliers of safety-critical embedded products to develop more cost and energy-aware solutions. To ensure this impact, a community will be created around the project. SAFURE comprises 7 industrial manufacturers, 4 leading universities and research centres and 1 SME.

[H2020 CPS 예비후보05]

Testing Cyber-Physical Systems under Uncertainty: Systematic, Extensible, and Configurable Model-based and Search-based Testing Methodologies (U-Test)

http://cordis.europa.eu/project/rcn/194326_en.html

http://cordis.europa.eu/result/rcn/186496_en.html

Project ID: 645463

From 2015-01-01 to 2017-12-31, ongoing project

Total cost: EUR 3 713 233,75 (EU contribution: EUR 3 713 233,75)

Coordinated in: Norway OSLO MEDTECH FORENING

Funded under: H2020-EU.2.1.1.1. - A new generation of components and systems: Engineering of advanced embedded and energy and resource efficient components and systems

Topic(s): ICT-01-2014 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2014-1 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

Uncertainty is intrinsic in Cyber-Physical Systems (CPSs) due to novel interactions of embedded systems, networking equipment, cloud infrastructures, and humans. CPSs have become predominant in critical domains and necessitate the implementation of proper mechanisms to deal with uncertainty during their operation at an acceptable cost avoiding unwarranted threats to its users and environment. One way to guarantee the correct implementation of such mechanisms is via **automated and systematic Model-Based Testing (MBT)**—a way of improving **dependability**.

U-Test will improve the dependability of CPSs by defining extensible MBT frameworks supporting holistic MBT of CPSs under uncertainty in a cost-effective manner. More specifically our objectives are: 1) Provide a comprehensive and extensible taxonomy of uncertainties classifying uncertainties, their properties, and relationships; 2) An Uncertainty Modelling Framework (UMF) to support modelling uncertainties at various levels relying on exiting modelling/testing standards; 3) Defining an intelligent way to evolve uncertainty models developed using UMF towards realistic unknown uncertainty models using search algorithms (e.g., Genetic Algorithms); 4) Generating cost-effective test cases from uncertainty and evolved models.

U-Test consortium encompasses domain experts from various facets of CPSs, i.e., software, embedded systems, distributed systems, and cloud infrastructure. We have chosen two case studies from diverse domains including Handling Systems and Geo Sports to assess the cost-effectiveness of U-Test. The solutions will be integrated into two key commercial tools available in the market: ModelBus/Fokus!MBT and CertifyIt. Moreover, the solutions will be deployed into the actual practise in addition to standardization to achieve a wider impact within Logistics, Geo Sports, and Healthcare domains and further facilitate interoperability among tools and technologies.

[H2020 CPS 예비후보06]

European Network of competencies and platforms for Enabling SME from any sector building Innovative CPS products to sustain demand for European manufacturing (EuroCPS)

http://cordis.europa.eu/project/rcn/194150_en.html

http://cordis.europa.eu/result/rcn/190410_en.html

Project ID: 644090

From 2015-02-01 to 2018-01-31, ongoing project

Total cost: EUR 9 127 173,01 (EU contribution: EUR 8 186 834,01)

Coordinated in: France

COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES

Funded under: H2020-EU.2.1.1.1. - A new generation of components and systems:

Engineering of advanced embedded and energy and resource efficient components and systems

Topic(s): ICT-01-2014 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2014-1See other projects for this call

Funding scheme: IA - Innovation action

Objective

SMEs play a key role in European economies; they constitute the largest business block and provide the bulk of employment. They generate most of the innovative ideas for ICT and CPS-enabled IoT products, two areas which represent an inflection point for innovators and industry in Europe.

EuroCPS is an ambitious project that aims to arm Europe with a network of design centres that will boost and initiate synergies between SMEs, major CPS-platforms, and CPS-competency providers, to capture the emerging markets for IoT products.

EuroCPS will:

1. Leverage the existing regional ecosystem across the full value chain (from micro-electronics, smart systems, and CPS, to high value added products and services) and range of expertise and competencies to provide innovators from any sector with an easy path to build innovative CPS-enabled systems.
2. Act as a one-stop-shop that provides a critical mass of technologies and competencies by facilitating user-supplier partnerships across value-chains and regions. Hence the typical development time of innovative for CPS applications will be significantly decreased through the ease of access to the platforms, and coaching by the competence partners within EuroCPS.
3. Link software, system and nano-electronic industries along the full CPS value chain to demonstrate a new cooperation model evidenced by experiments initiated and led by innovators that translate the rich pool of ideas from end users into implementation of CPS applications made in Europe.

EuroCPS gathers major European system suppliers and world class research centres and technology providers, all rooted in the top European regional ecosystems. Based on strong foundations in European and national initiatives, EuroCPS will, through pan European collaboration and knowledge exchange and access to the strong value chain in this strategic sector, significantly reduce development time and certification efforts, thus putting Europe at the cutting edge of CPS development and implementation.

[H2020 CPS 예비후보07]

Trusted Apps for open CPS (TAPPS)

http://cordis.europa.eu/project/rcn/194284_en.html

http://cordis.europa.eu/result/rcn/192883_en.html

Project ID: 645119

From 2015-01-01 to 2017-12-31, ongoing project

Total cost: EUR 3 885 484 (EU contribution: EUR 3 885 484)

Coordinated in: Germany FORTISS GMBH

Funded under: H2020-EU.2.1.1.1. - A new generation of components and systems: Engineering of advanced embedded and energy and resource efficient components and systems

Topic(s): ICT-01-2014 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2014-1 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

Open and smart cyber-physical systems (CPS) are considered to be the next revolution in ICT with enormous economic potential enabling novel business models for integrated services and products. In many areas of CPS devices, there is a strong trend towards open systems, which can be extended during operation instantly adding functionalities on demand. The main goal of the TAPPS (Trusted Apps for open CPS) project is the development of **a platform for CPS Apps**, which can also access and modify device internals. Therefore, the solution will address all necessary layers from hardware over software to an app store concept always ensuring security and full real-time support for the applications. The extensibility and the pervasive trusted environment of TAPPS are important differentiators that will enable new market extensions to keep pace with user expectations and latest technology.

As current, rich execution platforms for apps are limited in security, the project will develop a parallel, real-time Trusted Execution Environment (TEE) for highly-trusted CPS Apps. The TEE is located separately from existing the execution environment inside the System Control Units and exploits functionalities provided by the novel hardware-, processor- and network-centric security mechanisms as well as a hypervisor for virtualization. Furthermore, TAPPS will provide and validate an end-to-end solution for development and deployment of trusted apps, including an App Store and a model-based tool chain for trusted application development including verification tools. This multi-level trusted Apps platform and tool chain are matured and validated in health and automotive application domains using industrial, realistic use cases paving the way for future exploitation in further demanding application domains.

[H2020 CPS 예비후보08]

Integrated Modelling, Fault Management, Verification and Reliable Design Environment for Cyber-Physical Systems (IMMORTAL)

http://cordis.europa.eu/project/rcn/194260_en.html

http://cordis.europa.eu/result/rcn/190163_en.html

Project ID: 644905

From 2015-03-01 to 2018-02-28, ongoing project

Total cost: EUR 3 996 652,50 (EU contribution: EUR 3 996 652,50)

Coordinated in: Estonia TALLINNA TEHNIKAULIKOOL

Funded under: H2020-EU.2.1.1.1. - A new generation of components and systems:

Engineering of advanced embedded and energy and resource efficient components and systems

Topic(s): ICT-01-2014 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2014-1 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

In IMMORTAL, a consortium of leading European academic and industrial players aim at combining their expertise in developing an integrated, cross-layer modelling based tool framework for **fault management, verification and reliable design** of dependable Cyber-Physical Systems (CPS).

Recently, the world has seen emerging CPS modelling frameworks addressing various design aspects such as control, security, verification and validation. However, there have been no considerations for reliability and automated debug aspects of verification. The main aim is to fill this gap by introducing reliable design and automated system debug into CPS modelling. To reach this aim, the project will develop a cross-layer CPS model spanning device (analogue and digital), circuit, network architecture, firmware and software layers. In addition, a holistic fault model for fundamentally different error sources in CPSs (design bugs, wear-out and environmental effects) in a uniform manner will be proposed. Moreover, IMMORTAL plans to develop fault management infrastructure on top of the reliable design framework that would allow ultra-fast fault detection, isolation and recovery in the emerging many-core based CPS networked architectures that are expected to be increasingly adopted in the coming years.

As a result, the project will enable development of **dependable CPSs** with improved reliability and extended effective life-time, ageing and process variations. In line with the expected impacts of the Call, the project will have a significant impact in development time as well as maintenance costs of dependable cyber-physical systems.

The tool framework to be developed will be evaluated on a clearly specified real-world use-case of a satellite on-board-computer. However, since the results are more general and applicable to many application domains, including avionics, automotive and telecommunication, demonstration of the framework tools will be applied to CPS examples from other domains as well.

[H2020 CPS 예비후보09]

Agile, eXtensible, fast I/O Module for the cyber-physical era (AXIOM)

http://cordis.europa.eu/project/rcn/194328_en.html

http://cordis.europa.eu/result/rcn/186489_en.html

Project ID: 645496

From 2015-02-01 to 2018-01-31, ongoing project

Total cost: EUR 3 945 937,50 (EU contribution: EUR 3 945 937,50)

Coordinated in: Italy UNIVERSITA' DEGLI STUDI DI SIENA

Funded under: H2020-EU.2.1.1.1. - A new generation of components and systems:

Engineering of advanced embedded and energy and resource efficient components and systems

Topic(s): ICT-01-2014 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2014-1 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

We are entering the Cyber-Physical age, in which both objects and people will become nodes of the same digital network for exchanging information. Therefore, in our imaginary, the general expectation is that “things” or systems will become somewhat smart as people, allowing rapid and close interactions not only system-system, but also human-system, system-human. More scientifically, we expect that such Cyber-Physical Systems (CPSs) will at least react in **real-time**, have enough **computational power** for the assigned tasks, consume the least possible energy for such task (**energy efficiency**), scale up through **modularity**, allow for an **easy programmability** across performance scaling and exploit at best existing standards at **minimal costs**. The whole set of these expectations impose scientific and technological challenges that need to be properly addressed.

The AXIOM project (Agile, eXtensible, fast I/O Module) aims at **researching new software/hardware architectures for CPSs** to meet the above expectations. The technical approach aims at solving fundamental problems to enable easy programmability of multi-core multi-board systems through the **open-source OmpSs** programming model, leveraging **Distributed Shared Memory (DSM)** inspired concepts across the modules. The OmpSs will allow accelerating functions through an **FPGA (Agility)**. In particular, to the best of our knowledge, this is the first time that DSM will be effectively demonstrated on an **embedded modular system (eXtensibility)**. Modular scalability will be possible thanks to a fast interconnect that will enrich the module. To this aim, an innovative modular **ARM-based** board with enhanced capabilities for interfacing with the physical world will be designed and demonstrated in key scenarios such as Smart Video-Surveillance and Smart Living/Home (Domotic).

[H2020 CPS 예비후보10]

Towards Cyber-Physical Systems Engineering Tools Interoperability Standardisation (CP-SETIS)
표준화

http://cordis.europa.eu/project/rcn/194289_en.html

http://cordis.europa.eu/result/rcn/190233_en.html

Project ID: 645149

From 2015-03-01 to 2017-05-31, closed project

Total cost: EUR 780 000,37 (EU contribution: EUR 698 895)

Coordinated in: Germany

Funded under: H2020-EU.2.1.1.1. - A new generation of components and systems:
Engineering of advanced embedded and energy and resource efficient components and systems

Topic(s): ICT-01-2014 - Smart Cyber-Physical Systems

Call for proposal: H2020-ICT-2014-1See other projects for this call

Funding scheme: IA - Innovation action

Objective

CPS require multiple engineering competences across various engineering disciplines. The development of such systems is a huge challenge, also because of the heterogeneity of engineering tools involved in development platforms across the development lifecycle. In order to overcome this challenge, past and on-going EU research projects have developed the basis for an International Open Standard for Development Tool Interoperability, the so called **Interoperability Specification (IOS)**. However, due to lack of coordination, current IOS related activities, especially with respect to its standardization and possible extensions, are un-coordinated, endangering the huge financial effort that has been put into the IOS and the chance to establish it as a formal Standard.

The main goal of CP-SETIS is **to conceive and set up a sustainable organizational structure as a platform joining all stakeholders**, to coordinate all IOS related activities, especially the formal standardization and further extensions of the IOS. This organizational structure will be implemented within existing, sustainable European organizations, like the **ARTEMIS-IA Working Groups** or similar, that are open to all stakeholders.

CP-SETIS will ensure the support of all stakeholders for this structure, its operational rules, its implementation within existing structures and, most important, their commitment to coordinate all IOS related activities within this structure.

In this way, CP-SETIS will secure the huge effort – both in terms of manpower as well as in terms of financial support – that has been put into the IOS, furthering and enabling the setup of the IOS as a formal standard, and enabling the enormous innovation potential of the IOS both, for innovations in interoperable tools with new functionalities needed for future generations of Cyber-Physical Systems, and for innovations in future CPS themselves.

CP-SETIS will also use lessons learned during this process to update Standardization Research Agendas.

<부록 18> H2020 빅데이터 예비 후보

[H2020 빅데이터 예비후보1]

Perceptive Sentinel – BIG DATA knowledge extraction and re-creation platform
(PerceptiveSentinel)

http://cordis.europa.eu/project/rcn/212428_en.html

Project ID: 776115

From 2018-01-01 to 2020-06-30, Grant Agreement signed

Total cost: EUR 1,983,187 (EU contribution: EUR 1,983,187)

Funded under: H2020-EU.2.1.6.1.2. - Boost innovation between space and non-space sectors

H2020-EU.2.1.6.3. - Enabling exploitation of space data

Topic(s): EO-2-2017 - EO Big Data Shift

Call for proposal: H2020-EO-2017 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

PerceptiveSentinel project will deliver **PerceptiveSentinel platform - an INTERMEDIATE EO service** that will through seamless access to highly pre-processed DATA provide:

- MODELLING and PUBLISHING capabilities for
- DESIGN, EXPOSURE and EXPLOITATION of EO-processing chains for
- FORECASTING, MONITORING and HISTORICAL ANALYSIS in different domain fields, based on
- MULTI-TEMPORAL and MULTI-SPECTRAL EO-Data and Non-EO Data MODELLING.

PerceptiveSentinel platform will combine BIG-DATA SOURCES, (1) both spatial and non-spatial, (2) whether remote sensed, observed in-situ or derived (3) whether public, private or proprietary—onto a single system. The data will be transformed into action by applying STREAMING MACHINE LEARNING to UNLOCK its value.

Platform's capabilities will significantly shorten development cycles, enabling rapid development of EO services, opening new opportunities for large-scale exploitation of data and provide the means for new EO "downstream players" to smoothly enter the market without the investment into their own storage and processing capabilities.

Commercialisation approach will be based on (1) FREE and OPEN access to services and data, (2) OPEN-VALUE-CHAIN approach, (3) LOW COST of services and on (4) DELIVERING VALUE TO END USERS.

It is our intention that PerceptiveSentinel platform becomes FIRST value and benefit sharing EO platform in the world. We intent to expose an alternative business model building on COOPERATIVE VALUE CREATION and capture, including all aspects of EO value chain from data providing, product/service development, publishing and consuming. We believe in synergetic effects of cooperation, using PerceptiveSentinel platform to establish alliances with external partners, taking an open approach to participation and building joint commercial success on the delivery of new added value.

[H2020 빅 데이터 예비후보2]

API-ecosystem for cross-sectorial exchange of 3D personal data (BodyPass)

http://cordis.europa.eu/project/rcn/212483_en.html

Project ID: 779780

From 2018-01-01 to 2020-12-31, Grant Agreement signed

Total cost: EUR 3,110,581.25 (EU contribution: EUR 2,552,018.38)

Topic(s): ICT-14-2016-2017 - Big Data PPP: cross-sectorial and cross-lingual data integration and experimentation

LSE CORP SRL Italy (EU contribution: EUR 115,704.75) VIA TORQUATO TASSO 7 20123 MILANO MI Italy

Activity type: Private for-profit entities (excluding Higher or Secondary Education Establishments)
Contact the organisation

BodyPass aims to break barriers between **health sector** and **consumer goods sector** and eliminate the current data silos. The main objective of BodyPass is to foster exchange, linking and re-use, as well as to integrate **3D data assets** from the two sectors. For this, BodyPass has to adapt and create tools that allow a secure exchange of information between data owners, companies and subjects (patients and customers).

3D personal data is type of data that contains useful information for product design, online sale services, medical research and patient follow-up.

Currently **hospitals store and grow massive collections of 3D data** that are not accessible by researchers, professionals and companies. About 2.7 petabytes a year stored in the EU26.

In parallel to the advances made in the health sector, new 3D body-surface scanning technology has been developed for the goods consumer sector, namely apparel, animation and art.

Moreover, new low-cost scanning technologies are expected to exponentially increase 3D data creation. It is estimated that currently one person is scanned every 15 minutes in the US and Europe. And increasing.

The 3D data of the health sector contains the body shape information, not only internal body information. These data could be used by designers and manufacturers of the consumer goods sector. At the same time, although 3D body-surface scanners have been developed primarily for the clothing industry, 3D scanners' low cost, non-invasive character, and ease of use make them appealing for widespread clinical applications and large-scale epidemiological surveys.

However, companies and professionals of the consumer goods sector cannot access the 3D data of health sector. And vice versa. Even exchanging information between data owners in the same sector is a big problem today. It is necessary to overcome problems related with data privacy and the processing of huge 3D datasets.

[H2020 빅 데이터 예비후보3]

Data-driven decision making for a more efficient society (BiggerDecisions)

http://cordis.europa.eu/project/rcn/211570_en.html

Project ID: 778615

From 2017-09-01 to 2019-02-28, ongoing project

Total cost: EUR 1,830,625 (EU contribution: EUR 1,281,437)

Topic(s): SMEInst-01-2016-2017 - Open Disruptive Innovation Scheme

Objective

The rapid diffusion of digital information technologies has created an explosive growth in the amount of digital data. 90% of the world's data today has been created in the last two years alone. The availability of this data has opened up an opportunity to create significant value on all levels of the society. More extensive use of this data for data-driven decision-making (by companies, governments, researchers) could improve European GDP by more than €120 billion by 2020. However, this opportunity is currently restricted by privacy regulations, distrust and business secrets.

Sharemind is the first technology that truly enables the use of confidential data for data-driven decision-making, and thereby fundamentally enhances the potential value of big data. The novelty of Sharemind lies in its **ability to extract useful information from private data without even seeing the data**. Sharemind makes new types and more data available for sharing/analysis and simplifies the data analysis process, and thereby enables applications that have been impossible before. Users of Sharemind gain new insights that can be used for making decisions, improving productivity, optimising marketing, developing new products and services and implementing new business models.

Sharemind has been validated in an operational environment with five piloting customers (TRL7). Our go-to-market action plan foresees approaching our existing list of interested prospects as soon as **the limitations to Sharemind global scale-up** have been removed in the **BiggerDecisions** project. We have identified 250 customer leads, of which 30 are currently in sales negotiations with us.

One of the main factors supporting our commercialization efforts is the enforcement of General Data Protection Regulation in 2018, requiring a level of data privacy and security that is beyond what most organisations are currently equipped to manage.

We will execute a business strategy yielding over €94 million in revenues in 2021.

Expand / Contract(-)

[H2020 빅 데이터 예비후보4]

MAKING Sustainable development and WELL-being frameworks work for policy analysis (MAKSWELL)

http://cordis.europa.eu/project/rcn/212388_en.html

Project ID: 770643

From 2017-11-01 to 2020-04-30, ongoing project

Total cost: EUR 934,296.88 (EU contribution: EUR 934,296.88)

Funded under: H2020-EU.3.6.1.1. - The mechanisms to promote smart, sustainable and inclusive growth H2020-EU.3.6.2. - Innovative societies

Topic(s): CO-CREATION-07-2017 - Towards a new growth strategy in Europe - Improved economic and social measurement, data and official statistics

Call for proposal: H2020-SC6-CO-CREATION-2017 See other projects for this call

Funding scheme: CSA - Coordination and support action

HUNGARIAN CENTRAL STATISTICAL OFFICE Hungary (EU contribution: EUR 43,345)

Activity type: Public bodies (excluding Research Organisations and Secondary or Higher Education Establishments)

CONSORZIO PER LO SVILUPPO DELLE METODOLOGIE E DELLE INNOVAZIONI NELLE PUBBLICHE AMMINISTRAZIONI Italy (EU contribution: EUR 88,500)

MAKSWELL project proposes to extend and harmonising the indicators able to capture the main characteristics of the **beyond-GDP approach** proposing a new framework that includes them in the evaluation of the public policies. The main goals of the project can be summarized in three main objectives. In particular, the project aims at:

1_ **building up a database** for a wide set of EU countries that select and harmonize the national framework on well-being as well as the available SDG indicators;

2_ **improving the Database** both in relation to the timeliness and to the integration with big data measures and the methodologies able to reach these extensions. Particularly WP2 will extend the actual set of information available on well-being and sustainability to including coherent new data sources (eg big data) able to derive coherent indicators for local analysis and disaggregation for other domain;

3_ **using the extended Database** for policy evaluation and built up a national pilot studies.

MAKSWELL project aims at creating a shared knowledge on the state of the art on relevant dimensions of sustainable development and on vulnerabilities and potentialities of society; on the most appropriate traditional and new data collection tools and modern statistical methods to have timely and accurate data.

The analysis and summary of the main features of the piloting process will produce recommendations. They will be substantially suggestions and best practices collections to maintain and update the prototype fed by these new and traditional data sources for more focused policy decisions. The knowledge will be built up through the contact and reciprocal fertilization of NSIs, academics and stakeholders, identifying and contacting key practitioners to participate in a virtual Working Group. projects results will also path ways for common research under FP9.

Expand / Contract(-)Collapse all

Participants

[H2020 빅 데이터 예비후보5]

MedAS: a Machine learning enabled Clinical Decision Support System to prevent prescription errors and improve patient safety (MEDAS)

http://cordis.europa.eu/project/rcn/210413_en.html

Project ID: 775047

From 2017-05-01 to 2017-08-31, closed project | MEDAS Website

Total cost: EUR 71,429 (EU contribution: EUR 50,000)

Coordinated in: Israel

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

H2020-EU.2.3.1. - Mainstreaming SME support, especially through a dedicated instrument

Topic(s): SMEInst-01-2016-2017 - Open Disruptive Innovation Scheme

Call for proposal: H2020-SMEINST-1-2016-2017 See other projects for this call

Funding scheme: SME-1 - SME instrument phase 1

Objective

Medication errors are the most common cause of adverse events in medication practice, although they are preventable. Only in Europe, prescription error rates range from 7.5% to 9.1% of the total managed medicines, representing a major public health issue, as some of those can even be fatal. They also represent a great economic burden to healthcare systems, with annual costs reaching €4.5B to €21.8B, depending on the country. MedAware's founders decided to take action on the theme in 2012, when they found out that a nine-year-old boy died simply because his primary care physician accidentally selected the wrong drug, on his electronic prescribing pull-down list. By realising that current solutions completely failed to save this boy, the team started developing several proof-of-concept algorithms that became our first prototype and then MedAS (MedAware Alerting System). MedAS is an innovative patient-specific Clinical Decision Support System, that identifies and alerts on prescription errors in real-time, and with greater than 80% accuracy. It utilizes big-data analytics and advanced machine learning to identify statistical outliers and to generate precise alerts that would otherwise be missed by existing CDSS. MedAS's effectiveness has already been proven both in retrospective trials and in real medical facilities in Israel and the USA: we improved patient safety, outcomes, and experience while dramatically reducing healthcare costs. Enabled by our next generation technology and given MedAS's unique capabilities, we now intend to build a solid technology platform and to deploy it to the European market. The main objective of the feasibility study is to assess MedAS from technical, commercial and financial perspectives. We will seize the opportunity to enter the constantly growing **CDSS market** (€51.8M in Europe by 2018) with our novel technology and platform, and estimate revenues of €22.1M by 2024, with a ROI of 8.9.

[H2020 빅 데이터 예비후보6]

Scalable Policy-aware linked data architecture for privacy, transparency and compliance (SPECIAL)

http://cordis.europa.eu/project/rcn/206343_en.html

Project ID: 731601

From 2017-01-01 to 2019-12-31, ongoing project

Total cost: EUR 3,991,388.75 (EU contribution: EUR 3,991,388.75)

Coordinated in: France

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-18-2016 - Big data PPP: privacy-preserving big data technologies

Call for proposal: H2020-ICT-2016-1 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

The SPECIAL project will address the contradiction between Big Data innovation and privacy-aware data protection by proposing a technical solution that makes both of these goals realistic. We will develop technology that: (i) **supports the acquisition of user consent** at collection time and the recording of both data and metadata (consent, policies, event data, context) according to legislative and user-specified policies; (ii) **caters for privacy-aware, secure workflows** that include usage/access control, transparency and compliance verification; (iii) **demonstrates robustness** in terms of performance, scalability and security all of which are necessary to support privacy preserving innovation in Big Data environments; and (iv) **provides a dashboard** with feedback and control features that make privacy in Big Data comprehensible and manageable for data subjects, controllers, and processors. SPECIAL shall allow citizens and organisations to share more data, while guaranteeing data protection compliance, thus enabling both trust and the creation of valuable new insights from shared data. Our vision will be realised and validated via real world use cases that - in order to be viable - need to overcome current challenges concerning the processing and sharing of data in a privacy preserving manner. In order to realise this vision, we will combine and significantly extend big data architectures to handle Linked Data, harness them with sticky policies as well as scalable queryable encryption, and develop advanced user interaction and control features: SPECIAL will build on top of the Big Data Europe and PrimeLife Projects, exploit their results, and further advance the state of the art of privacy enhancing technologies.

[H2020 빅 데이터 예비후보7]

Scalable Oblivious Data Analytics (SODA)

http://cordis.europa.eu/project/rcn/205932_en.html

Project ID: 731583

From 2017-01-01 to 2019-12-31, ongoing project

Total cost: EUR 2,980,610 (EU contribution: EUR 2,980,609.75)

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-18-2016 - Big data PPP: privacy-preserving big data technologies

Call for proposal: H2020-ICT-2016-1See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

More and more data is being generated, and analyzing this data drives knowledge and value creation across society. Unlocking this potential requires sharing of (often personal) data between organizations, but this meets unwillingness from data subjects and data controllers alike. Hence, techniques that protect personal information for data access, processing, and analysis are needed. To address this, the SODA project will enable **practical privacy-preserving analytics of information from multiple data assets using multi-party computation (MPC) techniques**. For this data does not need to be shared, only made available for encrypted processing. The main technological challenge is to make MPC scale to big data, where we will achieve substantial performance improvements. We embed MPC into a comprehensive privacy approach, demonstrated in an ICT-14.b and a **healthcare use case**.

Our first objective is to enable MPC for big data applications by scaling the performance. We follow a use case-driven approach, combining expertise from the domains of MPC and data analytics. Our second objective is to combine these improvements with a multidisciplinary approach towards privacy. By enabling differential privacy in the MPC setting aggregated results will not leak individual personal data. Legal analysis performed in a feedback loop with technical development will ensure improved compliance with EU data privacy regulation. User studies performed in a feedback loop with our consent control component will make data subjects more confident to have their data processed with our techniques. Our final objective is to validate our approach, by applying our results in a medical demonstrator originating from Philips practice and in a use case arising from the ICT-14.b data experimentation incubators. The techniques will be subjected to **public hacking challenges**. The technical innovations will be released as **open-source** improvements to the **FRESCO MPC framework**.

[H2020 빅 데이터 예비후보8]

BIG DATA FOR SMART SOCIETY (GATE)

http://cordis.europa.eu/project/rcn/211425_en.html

Project ID: 763566

From 2017-09-01 to 2018-08-31, ongoing project

Total cost: EUR 387,875 (EU contribution: EUR 387,875)

Funded under: H2020-EU.4.a. - Teaming of excellent research institutions and low performing RDI regions

Topic(s): WIDESPREAD-04-2017 - Teaming Phase 1

Call for proposal: H2020-WIDESPREAD-04-2017-TeamingPhase1See other projects for this call

Funding scheme: CSA - Coordination and support action

Objective

The general objective of GATE is to establish **Centre of Excellence on "Big Data for Smart Society – GATE"** that will fulfill the vision of Open Innovation through building a sustainable University-Government-Industry-Society ecosystem. The CoE will be established as joint initiative between Sofia University – the most prestigious educational and scientific hub in Bulgaria and Chalmers University of Technology, Sweden – leading European institution with extensive experience in research, education and innovation. CoE main research objective is **to advance the state-of-the-art in the whole Big Data Value Chain**, including development of advanced methods and tools for data collection from variety of structured and unstructured sources, data consistency checking and cleaning, data aggregation and linking, data processing, modeling and analysis, data delivery by providing both accessibility and proper visualization. Following challenges of the Programme Horizon 2020 and the Bulgarian Innovation Strategy for Smart Specialization 2014-2020, the project team selected the most promising Data Driven Innovation Pillars: Data Driven Government (Public Services based on Open Data); Data Driven Industry (Manufacturing and Production); Data Driven Society (Smart Cities); and Data Driven Science.

The specific objectives of Teaming 1 phase are to establish a solid background for the creation and sustainability of a CoE at a national, regional and EU level through (1) Elaboration of a detailed and robust business plan with a long term vision for setting-up and sustaining of the CoE; (2) Strengthening the research capacity and potential in Big Data; (3) Establishment of an international collaborative network of Big Data and related fields researchers; (4) Increasing quality of education and training and offering measures for motivation and involvement of the next-generation Early-Stage Researchers; and (5) Wide dissemination and promotion of project aims, activities and expected outputs.

<부록 19> H2020 HPC 예비 후보

[H2020 HPC 예비후보01]

Co-designed Innovation and System for Resilient Exascale Computing in Europe: From Applications to Silicon (**EuroEXA**)

http://cordis.europa.eu/project/rcn/210095_en.html

Project ID: 754337

From 2017-09-01 to 2021-02-28, ongoing project

Total cost: EUR 19,949,022.50 (EU contribution: EUR 19,949,022.50)

Funded under: H2020-EU.1.2.2. - FET Proactive

Topic(s): FETHPC-01-2016 - Co-design of HPC systems and applications

Call for proposal: H2020-FETHPC-2016 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

To achieve the demands of extreme scale and the delivery of exascale, we embrace the computing platform as a whole, not just component optimization or fault resilience. EuroEXA brings a holistic foundation from multiple European HPC projects and partners together with the industrial SME focus of MAX for FPGA data-flow; ICE for infrastructure; ALLIN for HPC tooling and ZPT to collapse the memory bottleneck; to co-design a ground-breaking platform capable of scaling peak performance to **400 PFLOP** in a peak system power envelope of **30MW**; over four times the performance at four times the energy efficiency of today's HPC platforms. Further, we target a PUE parity rating of 1.0 through use of renewables and immersion-based cooling.

We co-design a balanced architecture for both compute- and data-intensive applications using a cost-efficient, modular-integration approach enabled by novel inter-die links and the tape-out of a resulting **EuroEXA processing unit** with integration of FPGA for data-flow acceleration. We provide a homogenised software platform offering heterogeneous acceleration with scalable shared memory access and create a unique hybrid geographically-addressed, switching and topology interconnect within the rack while enabling the adoption of low-cost Ethernet switches offering low-Latency and high-switching bandwidth.

Working together with a rich mix of key HPC applications from across climate/weather, physics/energy and life-science/bioinformatics domains we will demonstrate the results of the project through the deployment of an integrated and operational peta-flop level prototype hosted at STFC. Supported by run-to-completion platform-wide resilience mechanisms, components will manage local failures, while communicating with higher levels of the stack. Monitored and controlled by advanced runtime capabilities, EuroEXA will demonstrate its co-design solution supporting both existing pre-exascale and project-developed exascale applications.

[H2020 HPC 예비후보02]

DEEP - Extreme Scale Technologies (**DEEP-EST**)

http://cordis.europa.eu/project/rcn/210094_en.html

Project ID: 754304

From 2017-07-01 to 2020-06-30, ongoing project

Total cost: EUR 15 873 341,25 (EU contribution: EUR 14 998 342)

Funded under: H2020-EU.1.2.2. - FET Proactive

Topic(s): FETHPC-01-2016 - Co-design of HPC systems and applications

Call for proposal: H2020-FETHPC-2016 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

The DEEP-EST ("DEEP - Extreme Scale Technologies") project will create a first incarnation of **the Modular Supercomputer Architecture (MSA)** and demonstrate its benefits. In the spirit of the DEEP and DEEP-ER projects, the MSA integrates compute modules with different performance characteristics into a single heterogeneous system. Each module is a parallel, clustered system of potentially large size. A federated network connects the module-specific interconnects. MSA brings substantial benefits for heterogeneous applications/workflows: each part can be run on an exactly matching system, improving time to solution and energy use. This is ideal for supercomputer centres running heterogeneous application mixes (higher throughput and energy efficiency). It also offers valuable flexibility to the compute providers, allowing the set of modules and their respective size to be tailored to actual usage.

The DEEP-EST prototype will include three modules: **general purpose Cluster Module** and **Extreme Scale Booster** supporting the full range of HPC applications, and **Data Analytics Module** specifically designed for **high-performance data analytics (HPDA)** workloads. Proven programming models and APIs from HPC (**combining MPI and OmpSs**) and HPDA will be extended and combined with a significantly enhanced resource management and scheduling system to enable straightforward use of the new architecture and achieve highest system utilisation and performance. Scalability projections will be given up to the Exascale performance class. The **DEEP-EST prototype** will be defined in close co-design between applications, system software and system component architects. Its implementation will employ European integration, network and software technologies. Six ambitious and highly relevant European applications from HPC and HPDA domains will drive the co-design, serving to evaluate the DEEP EST prototype and demonstrate the benefits of its innovative Modular Supercomputer Architecture.

[H2020 HPC 예비후보03]

Transparent heterogeneous hardware Architecture deployment for eNergy Gain in Operation (**TANGO**)

http://cordis.europa.eu/project/rcn/200021_en.html

Project ID: 687584

From 2016-01-01 to 2018-12-31, ongoing project

Total cost: EUR 3 199 625 (EU contribution:)EUR 3 199 625

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-04-2015 - Customised and low power computing

Call for proposal: H2020-ICT-2015See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

Computer systems have faced significant power challenges over the past 20 years; these challenges have shifted from the devices and circuits level, to their current position as first-order constraints for system architects and software developers. TANGO's goal is **to characterise factors which affect power consumption in software development and operation for heterogeneous parallel hardware environments**. Our main contribution is the combination of requirements engineering and design modelling for self-adaptive software systems, with power consumption awareness in relation to these environments. The energy efficiency and application quality factors are integrated in the application lifecycle (design, implementation, operation). To support this, the key novelty of the project is a reference architecture and its implementation. Moreover, a programming model with built-in support for various hardware architectures including heterogeneous clusters, heterogeneous chips and programmable logic devices will be provided. TANGO will create a new cross-layer programming approach for heterogeneous parallel hardware architectures featuring automatic code generation including software and hardware modelling. This will consider power, performance, data location and time criticality optimization, in addition to security and dependability on the target hardware architecture. These results will be demonstrated in two real-world applications: reconfigurable power optimized connected platform and HPC. In order to improve collaboration and sustainability of TANGO's and fellow projects results, TANGO considers the foundation of a Research Alliance in which complementary research efforts into novel programming approaches will nucleate, leading to a strong research collaboration and effective integration of project results.

[H2020 HPC 예비후보04]

European Exascale System Interconnect and Storage (**ExaNeSt**)

http://cordis.europa.eu/project/rcn/197934_en.html

Project ID: 671553

From 2015-12-01 to 2018-11-30, ongoing project

Total cost: EUR 8 442 547,50 (EU contribution: EUR 8 442 547,50)

Funded under: H2020-EU.1.2.2. - FET Proactive

Topic(s): FETHPC-1-2014 - HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications

Call for proposal: H2020-FETHPC-2014 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

ExaNeSt will develop, evaluate, and prototype the physical platform and architectural solution for a unified Communication and Storage Interconnect and the physical rack and environmental structures required to deliver **European Exascale Systems**. The consortium brings technology, skills, and knowledge across the entire value chain from computing IP to packaging and system deployment; and from operating systems, storage, and communication to HPC with big data management, algorithms, applications, and frameworks. Building on a decade of advanced R&D, ExaNeSt will deliver the solution that can support exascale deployment in the follow-up industrial commercialization phases. Using direction from the **ETP4HPC roadmap** and soon-available high density and efficiency compute, we will model, simulate, and validate through prototype, a system with:

1. High throughput, low latency connectivity, suitable for exascale-level compute, their storage, and I/O, with congestion mitigation, QoS guarantees, and resilience.
2. Support for distributed storage located with the compute elements providing low latency that non-volatile memories require, while reducing energy, complexity, and costs.
3. Support for task-to-data sw locality models to ensure minimum data communication energy overheads and property maintenance in databases.
4. Hyper-density system integration scheme that will develop a modular, commercial, European-sourced advanced cooling system for exascale in ~200 racks while maintaining reliability and cost of ownership.
5. The platform management scheme for big-data I/O to this resilient, unified distributed storage compute architecture.
6. Demonstrate the applicability of the platform for the complete spectrum of Big Data applications, e.g. from HPC simulations to Business Intelligence support.

All aspects will be steered and validated with the first-hand experience of HPC applications and experts, through kernel turning and subsequent data management and application analysis.

[H2020 HPC 예비후보05]

Next Generation I/O for Exascale (**NEXTGenIO**)

http://cordis.europa.eu/project/rcn/197539_en.html

Project ID: 671591

From 2015-10-01 to 2018-09-30, ongoing project

Total cost: EUR 8 114 504 (EU contribution: EUR 8 114 504)

Funded under: H2020-EU.1.2.2. - FET Proactive

Topic(s): FETHPC-1-2014 - HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications

Call for proposal: H2020-FETHPC-2014

Funding scheme: RIA - Research and Innovation action

Objective

The overall objective of the Next Generation I/O Project (NEXTGenIO) is to design and prototype a new, scalable, high-performance, energy efficient computing platform designed to address the challenge of delivering scalable I/O performance to applications at the Exascale. It will achieve this using **highly innovative, non-volatile, dual in-line memory modules (NV-DIMMs)**. These hardware and systemware developments will be coupled to a co-design approach driven by the needs of some of today's most demanding HPC applications. By meeting this overall objective, NEXTGenIO will solve a key part of the Exascale challenge and enable HPC and Big Data applications to overcome the limitations of today's HPC I/O subsystems.

Today most high-end HPC systems employ data storage separate from the main system and the I/O subsystem often struggles to deal with the degree of parallelism present. As we move into the domain of extreme parallelism at the Exascale we need to address I/O if such systems are to deliver appropriate performance and efficiency for their application user communities.

The NEXTGenIO project will explore the use of NV-DIMMs and associated systemware developments through a co-design process with three 'end-user' partners: a high-end academic HPC service provider, a numerical weather forecasting service provider and a commercial on-demand HPC service provider. These partners will develop a set of I/O workload simulators to allow quantitative improvements in I/O performance to be directly measured on the new system in a variety of research configurations. Systemware software developed in the project will include performance analysis tools, improved job schedulers that take into account data locality and energy efficiency, optimised programming models, and APIs and drivers for optimal use of the new I/O hierarchy.

The project will deliver immediately exploitable hardware and software results and show how to deliver high performance I/O at the Exascale.

[H2020 HPC 예비후보06]

Energy-efficient Heterogeneous COmputing at exaSCALE (**ECOSCALE**)

http://cordis.europa.eu/project/rcn/197939_en.html

Project ID: 671632

From 2015-10-01 to 2018-09-30, ongoing project

Total cost: EUR 4 237 397,50 (EU contribution: EUR 4 237 397,50)

Coordinated in: Greece

Funded under: H2020-EU.1.2.2. - FET Proactive

Topic(s): FETHPC-1-2014 - HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications

Call for proposal: H2020-FETHPC-2014 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

In order to reach exascale performance current HPC servers need to be improved. Simple scaling is not a feasible solution due to the increasing utility costs and power consumption limitations. Apart from improvements in implementation technology, what is needed is to refine the HPC application development as well as the architecture of the future HPC systems.

ECOSCALE tackles this challenge by proposing a scalable programming environment and hardware architecture tailored to the characteristics and trends of current and future HPC applications, reducing significantly the data traffic as well as the energy consumption and delays. We first propose **a novel heterogeneous energy-efficient hierarchical architecture** and **a hybrid MPI+OpenCL programming environment and runtime system**. The proposed architecture, programming model and runtime system follows a hierarchical approach where the system is partitioned into multiple autonomous Workers (i.e. compute nodes). Workers are interconnected in a tree-like structure in order to form larger Partitioned Global Address Space (PGAS) partitions, which are further hierarchically interconnected via an MPI protocol.

Secondly, to further increase the energy efficiency of the system as well as its resilience, the Workers will employ reconfigurable accelerators that can perform coherent memory accesses in the virtual address space utilizing an IOMMU. The ECOSCALE architecture will support shared partitioned reconfigurable resources accessed by any Worker in a PGAS partition, and, more importantly, automated hardware synthesis of these resources from an OpenCL-based programming model.

We follow a co-design approach that spans a scalable HPC hardware platform, a middleware layer, a programming and a runtime environment as well as a high-level design environment for mapping applications onto the system. A proof of concept prototype and a simulator will be built in order to run two real-world HPC applications and several benchmarks.

[H2020 HPC 예비후보07]

MANGO: exploring Manycore Architectures for Next-GeneratiOn HPC systems

http://cordis.europa.eu/project/rcn/197942_en.html

Project ID: 671668

From 2015-10-01 to 2018-09-30, ongoing project | MANGO Website

Total cost: EUR 5 801 820 (EU contribution: EUR 5 801 820)

Funded under: H2020-EU.1.2.2. - FET Proactive

Topic(s): FETHPC-1-2014 - HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications

Call for proposal: H2020-FETHPC-2014

Funding scheme: RIA - Research and Innovation action

Objective

MANGO targets **to achieve extreme resource efficiency in future QoS-sensitive HPC** through ambitious cross-boundary architecture exploration for performance/power/predictability (PPP) based on the definition of new-generation high-performance, power-efficient, heterogeneous architectures with native mechanisms for isolation and quality-of-service, and an innovative two-phase passive cooling system. Its disruptive approach will involve many interrelated mechanisms at various architectural levels, including heterogeneous computing cores, memory architectures, interconnects, run-time resource management, power monitoring and cooling, to the programming models. The system architecture will be inherently heterogeneous as an enabler for efficiency and application-based customization, where general-purpose compute nodes (GN) are intertwined with heterogeneous acceleration nodes (HN), linked by an across-boundary homogeneous interconnect. It will provide guarantees for predictability, bandwidth and latency for the whole HN node infrastructure, allowing dynamic adaptation to applications. MANGO will develop a toolset for PPP and explore holistic pro-active thermal and power management for energy optimization including chip, board and rack cooling levels, creating a hitherto inexistent link between HW and SW effects at all layers. Project will build an effective large-scale emulation platform. The architecture will be validated through noticeable examples of application with QoS and high-performance requirements.

Ultimately, the combined interplay of the multi-level innovative solutions brought by MANGO will result in a new positioning in the PPP space, ensuring sustainable performance as high as **100 PFLOPS** for the realistic levels of power consumption (<15MWatt) delivered to QoS-sensitive applications in large-scale capacity computing scenarios providing essential building blocks at the architectural level enabling the full realization of the **ETP4HPC strategic research agenda**

[H2020 HPC 예비후보08]

European Exascale Processor Memory Node Design (**ExaNoDe**)

http://cordis.europa.eu/project/rcn/197936_en.html

Project ID: 671578

From 2015-10-01 to 2018-09-30, ongoing project

Total cost: EUR 8 629 247,50 (EU contribution:)EUR 8 629 247,50

Funded under: H2020-EU.1.2.2. - FET Proactive

Topic(s): FETHPC-1-2014 - HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications

Call for proposal: H2020-FETHPC-2014 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

ExaNoDe will investigate, develop integrate and validate the building blocks (technology readiness level 5) for a highly efficient, highly integrated, multi-way, high-performance, heterogeneous compute element aimed towards exascale computing. It will build on multiple European initiatives for scalable computing, utilizing low- power processors and advanced nanotechnologies. ExaNoDe will draw heavily on the Unimem memory and system design paradigm defined within the EUROSERVER FP7 project, providing low-latency, high-bandwidth and resilient memory access, scalable to Exabyte levels.

The ExaNoDe compute element aims towards exascale compute goals through:

- **Integration of the most advanced low-power processors and accelerators** (across scalar, SIMD, GPGPU and FPGA processing elements) supported by research and innovation in the deployment of associated nanotechnologies and in the mechanical requirements to enable the development of a high-density, high-performance integrated compute element with advanced thermal characteristics and connectivity to the next generation of system interconnect and storage;
- Undertaking essential research **to ensure the ExaNoDe compute element provides necessary support of HPC applications** including I/O and storage virtualization techniques, operating system and semantically aware runtime capabilities and PGAS, OpenMP and MPI paradigms;
- **The development of a hardware emulation of interconnect** to enable the evaluation of Unimem for the deployment of multiple compute elements and to leverage the potential of the ExaNoDe approach for HPC applications.

Each aspect of ExaNoDe is aligned with the goals of the **ETP4HPC**. The work will be steered by first-hand experience and analysis of high-performance applications and their requirements; investigations being carried out with “mini-application” abstractions and the tuning of their kernels.

[H2020 HPC 예비후보09]

Exploiting eXascale Technology with Reconfigurable Architectures (**EXTRA**)

http://cordis.europa.eu/project/rcn/197940_en.html

Project ID: 671653

From 2015-09-01 to 2018-08-31, ongoing project

Total cost: EUR 3 989 931,25 (EU contribution: EUR 3 989 930,50)

Funded under: H2020-EU.1.2.2. - FET Proactive

Topic(s): FETHPC-1-2014 - HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications

Call for proposal: H2020-FETHPC-2014 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

To handle the stringent performance requirements of future exascale High Performance Computing (HPC) applications, HPC systems need ultra-efficient heterogeneous compute nodes. To reduce power and increase performance, such compute nodes will require reconfiguration as an intrinsic feature, so that specific HPC application features can be optimally accelerated at all times, even if they regularly change over time.

In the EXTRA project, we create **a new and flexible exploration platform** for developing reconfigurable architectures, design tools and HPC applications with run-time reconfiguration built-in from the start. The idea is to enable the efficient co-design and joint optimization of architecture, tools, applications, and reconfiguration technology in order to prepare for the necessary HPC hardware nodes of the future.

The project EXTRA covers the complete chain from architecture up to the application:

- More coarse-grain reconfigurable architectures that allow reconfiguration on higher functionality levels and therefore provide much faster reconfiguration than at the bit level.
- The development of just-in time synthesis tools that are optimized for fast (but still efficient) re-synthesis of application phases to new, specialized implementations through reconfiguration.
- The optimization of applications that maximally exploit reconfiguration.
- Suggestions for improvements to reconfigurable technologies to enable the proposed reconfiguration of the architectures.

In conclusion, EXTRA focuses on **the fundamental building blocks for run-time reconfigurable exascale HPC systems**: new reconfigurable architectures with very low reconfiguration overhead, new tools that truly take reconfiguration as a design concept, and applications that are tuned to maximally exploit run-time reconfiguration techniques.

Our goal is to provide the European platform for run-time reconfiguration to maintain Europe's competitive edge and leadership in run-time reconfigurable computing.

[H2020 HPC 예비후보10]

Runtime Exploitation of Application Dynamism for Energy-efficient eXascale computing
(**READEX**)

http://cordis.europa.eu/project/rcn/197941_en.html

Project ID: 671657

From 2015-09-01 to 2018-08-31, ongoing project | READEX Website

Total cost: EUR 3 534 198,75 (EU contribution: EUR 3 534 198,75)

Funded under: H2020-EU.1.2.2. - FET Proactive

Topic(s): FETHPC-1-2014 - HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications

Call for proposal: H2020-FETHPC-2014

Funding scheme: RIA - Research and Innovation action

Objective

High Performance Computing (HPC) has become a major instrument for many scientific and industrial fields to generate new insights and product developments. There is a continuous demand for growing compute power, leading to a constant increase in system size and complexity. Efficiently utilizing the resources provided on Exascale systems will be a challenging task, potentially causing a large amount of underutilized resources and wasted energy. Parameters for adjusting the system to application requirements exist both on the hardware and on the system software level but are mostly unused today. Moreover, accelerators and co-processors offer a significant performance improvement at the cost of increased overhead, e.g., for data-transfers.

While HPC applications are usually highly compute intensive, they also exhibit a large degree of dynamic behaviour, e.g., the alternation between communication phases and compute kernels. Manually detecting and leveraging this dynamism to improve energy-efficiency is a tedious task that is commonly neglected by developers. However, using an automatic optimization approach, application dynamism can be detected at design-time and used to generate optimized system configurations. A light-weight run-time system will then detect this dynamic behaviour in production and switch parameter configurations if beneficial for the performance and energy-efficiency of the application. The READEX project will develop **an integrated tool-suite** and the READEX Programming Paradigm to exploit application domain knowledge, together achieving an improvement in energy-efficiency of up to 22.5%.

Driven by a consortium of European experts from academia, HPC resource providers, and industry, the READEX project will develop **a tools-aided methodology** to exploit the dynamic behaviour of applications to achieve improved energy-efficiency and performance. The developed tool-suite will be efficient and scalable to support current and future extreme scale systems.

<부록 20> H2020 클라우드 예비 후보

[H2020 클라우드 예비후보1]

Novel Enablers for Cloud Slicing (**NECOS**)

http://cordis.europa.eu/project/rcn/212036_en.html

Project ID: 777067

From 2017-11-01 to 2019-10-31, ongoing project

Total cost: EUR 1 494 906,25 (EU contribution: EUR 1 494 906,25)

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): EUB-01-2017 - Cloud Computing

Call for proposal: H2020-EUB-2017 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

The NECOS project addresses the limitations of current cloud computing infrastructures to respond to the demand of new services, as presented in **two use-cases** that will drive the whole execution of the project. The first use-case is **Telco service provider** focussed and is oriented towards the adoption of cloud computing in their large networks. The second use-case is targeting **the use of edge clouds** to support devices with low computation and storage capacity. The envisaged solution is based on a new concept – **Lightweight Slice Defined Cloud (LSDC)** – as an approach that extends the virtualization to all the resources in the involved networks and data centres and provides a uniform management with a high-level of orchestration.

The NECOS approach will be manifested in a platform whose main distinguishing features are:

1. The Slice as a Service -- a new deployment model. A slice is a grouping of resources managed as a whole, and that can accommodate service components, independent of other slices.
2. Embedded algorithms for an optimal allocation of resources to slices in the cloud and networking infrastructure, to respond to the dynamic changes of the various service demands.
3. A management and orchestration approach making use of artificial intelligence techniques in order to tackle with the complexity of large scale virtualized infrastructure environments.
4. Making reality the lightweight principle, in terms of low footprint components deployable on large number of small network and cloud devices at the edges of the network.

The NECOS platform will be based on state of the art **open software platforms**, which will be carefully selected, rather than start from scratch. This baseline platform will be enhanced with the management and orchestration algorithms and the APIs that will constitute the research activity of the project. Finally, the NECOS platform will be validated, in the context of the two proposed use cases, using the 5TONIC and FIBRE testing frameworks.

[H2020 클라우드 예비후보2]

Adaptive, Trustworthy, Manageable, Orchestrated, Secure, Privacy-assuring, Hybrid Ecosystem for REsilient Cloud Computing (**ATMOSPHERE**)

http://cordis.europa.eu/project/rcn/211963_en.html

Project ID: 777154

From 2017-11-01 to 2019-10-31, ongoing project

Total cost: EUR 1 499 375 (EU contribution: EUR 1 499 375)

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): EUB-01-2017 - Cloud Computing

Call for proposal: H2020-EUB-2017See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

ATMOSPHERE (Adaptive, Trustworthy, Manageable, Orchestrated, Secure Privacy-assuring Hybrid, Ecosystem for REsilient Cloud Computing) is a 24-month project aiming at the **design and development of an ecosystem** of a framework, platform and application of next generation trustworthy cloud services on top of an **intercontinental hybrid and federated resource pool**. The framework considers a broad spectrum of properties and their measures. The platform supports the building, deployment, measuring and evolution of trustworthy cloud resources, data network and data services. The platform is demonstrated on a sensitive scenario to build a cloud-enabled secure and trustworthy application related to distributed telemedicine.

[H2020 클라우드 예비후보3]

A Novel Framework For Multi-Cloud Services Development, Orchestration, Deployment And Continuous Management Fostering Cloud Technologies Uptake From Digital Smes And Startups (**UNICORN**)

http://cordis.europa.eu/project/rcn/206347_en.html

Project ID: 731846

From 2017-01-01 to 2019-12-31, ongoing project

Total cost: EUR 4 173 850 (EU contribution: EUR 3 415 487)

Coordinated in: Germany

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-06-2016 - Cloud Computing

Call for proposal: H2020-ICT-2016-1See other projects for this call

Funding scheme: IA - Innovation action

Objective

Unicorn aims **to simplify the design, deployment and management of secure and elastic –by design- multi-cloud services**. This will be achieved by a) development and design libraries that will provide security enforcement mechanisms, data privacy restrictions, monitoring metric collection and resource management; b) enabling continuous orchestration and automatic optimization of portable and dynamic cloud services running on virtual instances or micro-execution containers for increased security, data protection privacy and vast resource (de)-allocation. Towards this end, and building upon this umbrella concept, the Unicorn vision can be considered two-fold: to enable software developers to design and develop secure and elastic applications and to increase the awareness of all stakeholders, in particular SMEs and Startups.

Unicorn will reduce software release time and provide a powerful tool for **SMEs** to improve software design and continuous productivity enhancement. The innovation activities in the Unicorn project will be based upon existing solutions and developments to the largest possible extent, and build upon the S&T results of European RIA projects including CELAR, PaaSPort, PaaSWord and ARCADIA.

[H2020 클라우드 예비후보4]

ACTivating resource efficiency and large databases in the CLOUD (**ACTiCLOUD**)

http://cordis.europa.eu/project/rcn/206362_en.html

Project ID: 732366

From 2017-01-01 to 2019-12-31, ongoing project

Total cost: EUR 4 733 532,50 (EU contribution: EUR 4 733 532,50)

Coordinated in: Greece

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-06-2016 - Cloud Computing

Call for proposal: H2020-ICT-2016-1See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

Despite their proliferation as a dominant computing paradigm, cloud computing systems lack effective mechanisms to manage their vast amounts of resources efficiently, leading to severe resource waste and ultimately limiting their applicability to large classes of critical applications that pose non-moderate resource demands. This creates a significant need to lift existing technological barriers of actual fluidity and scalability of cloud resources towards promoting cloud computing as a critical cornerstone for digital economy.

ACTiCLOUD proposes a novel cloud computing architecture for drastically improved management of cloud resources, **targeting 1.5x increase in resource efficiency and more than 10x in scalability**. By utilizing modest investments on hardware intelligence that enables true resource disaggregation between multiple servers, we will progress current state-of-the-art in hypervisors and cloud management systems promoting holistic resource management at the rack scale and across distributed cloud sites. On top of this, we will evolve the ecosystem around in-memory databases, a core component for extremely demanding and highly critical classes of applications that up to now have faced severe difficulties in matching their resource requirements with state-of-the-art cloud offerings, with a final goal to provide **cost-efficient and highly performant DataBase-as-a-Service (DBaaS) cloud platforms**.

ACTiCLOUD builds on top of cutting-edge European technologies for cloud servers brought into the project by **Numascale** and **Kaleao**, and extends **OnApp's MicroVisor**, an innovative hypervisor to virtualize resources at the rack-scale. Furthermore it joins the forces of highly acclaimed academic institutions to address key research challenges and extend the capabilities of OpenStack and JVM. Finally, it applies the foreseen innovation to MonetDB, the column-store database pioneer, and Neo4j, the world-leader in graph databases.

[H2020 클라우드 예비후보5]

Cloud Orchestration at the Level of Application (**COLA**)

http://cordis.europa.eu/project/rcn/206078_en.html

Project ID: 731574

From 2017-01-01 to 2019-06-30, ongoing project

total cost: EUR 4 188 703,75 (EU contribution: EUR 2 975 081)

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-06-2016 - Cloud Computing

Call for proposal: H2020-ICT-2016-1 See other projects for this call

Funding scheme: IA - Innovation action

Objective

SMEs and public sector organizations increasingly investigate the possibilities to use cloud computing services in their everyday business conduct. Accessing services and resources in the cloud on-demand and in a flexible and elastic way could result in significant cost savings due to more efficient and convenient resource utilization that also replaces large investment costs with long term operational costs. On the other hand, the take up of cloud computing by **SMEs** and the **public sector** is still relatively low due to limited **application-level flexibility** and also **security** concerns. The Cloud Orchestration at the Level of Application (COLA) project aims **to increase the adoption of cloud computing services** by the above mentioned two strategic target communities. Typical industry and public sector applications require resource scalability and efficient resource utilization in order to serve a variable number of customers with dynamic resource demands, and to suitably optimize resource consumption and costs. However, the dynamic and intelligent utilization of cloud infrastructure resources from the perspective of cloud applications is not trivial. Although there have been several efforts to support the intelligent and coordinated deployment, and to a smaller extent also the run-time orchestration of cloud applications, no comprehensive solution has emerged until now that could be applied in large scale near operational level industry trials. The overall objective of the COLA project is that by building on and extending current research results, it will define and provide a reference implementation of a generic and pluggable framework that supports the optimal and secure deployment and run-time orchestration of cloud applications. COLA will demonstrate the applicability and impact of the solution via large scale near operational level SME and public sector pilots and demonstrators, and will also define a clear pathway how the innovation can be delivered to the market.

[H2020 클라우드 예비후보6]

DITAS: Data-intensive applications Improvement by moving daTA and computation in mixed cloud/fog environmentS (**DITAS**)

http://cordis.europa.eu/project/rcn/205979_en.html

Project ID: 731945

From 2017-01-01 to 2019-12-31, ongoing project

Total cost: EUR 4 890 066,25 (EU contribution: EUR 4 420 187,50)

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-06-2016 - Cloud Computing

Call for proposal: H2020-ICT-2016-1See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

There is an increasing need to develop data intensive applications able to manage more and more amounts of data coming from distributed and heterogeneous sources effectively, quickly, correctly, and securely. However, the current adoption of Cloud Computing paradigm is not fully appropriate to store and analyse such data: latency, security, and compliance are still significant barriers. At the same time, **Fog Computing** has emerged as a paradigm promising to fully exploit the potential of the edge of the network involving traditional devices as well as new generation of smart devices, which can process data closer to where they are produced and/or consumed but which cannot ensure the same reliability and scalability as cloud computing offers.

The goal of DITAS is **to propose a framework, composed by an SDK and an execution environment, which aims to overcome the barriers that now hamper the adoption of Cloud Computing and increase the adoption of Fog computing by exploiting the full potential of these two paradigms in a synergic way.** This will support the development and execution of data-intensive application that are now – and even more in the future – crucial for organizations and companies that want to manage their data in an efficient, reliable, scalable, and secure manner.

Abstractions provided in DITAS with **Data Virtualization** and Data Utility will expose the data to be managed by the application in terms of Virtual Data Containers which hide the complexity of the underlying infrastructure composed of heterogeneous data sources, smart devices, traditional servers, and sensor networks Distribution could also change dynamically. Conversely, Virtual Data Containers offer to developers the possibility to express requirements on data in terms of performance, quality, security and privacy thus to focus only on the application logic, leaving to the DITAS execution environment the responsibility of finding, processing, and delivering the data according to user needs

[H2020 클라우드 예비후보7]

Multi-cloud Execution-ware for Large-scale Optimized Data-Intensive Computing
(**MELODIC**)

http://cordis.europa.eu/project/rcn/206028_en.html

Project ID: 731664

From 2016-12-01 to 2019-11-30, ongoing project

Total cost: EUR 4 890 223,75 (EU contribution: EUR 4 890 223,75)

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-06-2016 - Cloud Computing

Call for proposal: H2020-ICT-2016-1 See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

MELODIC will enable **data-intensive applications to run within defined security, cost, and performance boundaries seamlessly on geographically distributed and federated cloud infrastructures**. Serving the user's needs and constraints, MELODIC will realise the potential of Cloud computing for big data and data-intensive applications by transparently taking advantage of distinct characteristics of available private and public clouds, dynamically optimise resource utilisation, consider data locality, conform to the user's privacy needs and service requirements, and counter vendor lock-in.

These benefits are achieved by integrating and extending the results and the open source platforms available from **three major European Cloud projects** with the Hadoop and Spark big data frameworks: The **PaaSage platform** will be used for intelligent and autonomic cross-cloud deployment and is extended with data aware modelling and deployment configuration reasoning; the **CACTOS platform** is extended with support for Hadoop and Spark in cross-Cloud application deployment and management; and the **PaaSword platform** will ensure unified data security and cross-Cloud privacy.

MELODIC will integrate with the existing open source development teams for these platforms and the contributions will be released back to the used platforms as open source. The integrated MELODIC platform will be maintained and exploited by a professional software house, and tested in several demanding real world applications: scalable Customer Relationship Management, real-time optimised traffic routing, and fast and scalable processing of genomic data.

[H2020 클라우드 예비후보8]

Cloud Large Scale Video Analysis (**Cloud-LSVA**)

http://cordis.europa.eu/project/rcn/199579_en.html

Project ID: 688099

From 2016-01-01 to 2018-12-31, ongoing project

Total cost: EUR 4 604 431,25 (EU contribution: EUR 4 604 431,25)

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-16-2015 - Big data - research

Call for proposal: H2020-ICT-2015See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

Cloud-LSVA will create **Big Data Technologies to address the open problem of a lack of software tools, and hardware platforms, to annotate petabyte scale video datasets**. The problem is of particular importance to the automotive industry. CMOS Image Sensors for Vehicles are the primary area of innovation for camera manufactures at present. They are the sensor that offers the most functionality for the price in a cost sensitive industry. By 2020 the typical mid-range car will have 10 cameras, be connected, and generate 10TB per day, without considering other sensors. Customer demand is for **Advanced Driver Assistance Systems (ADAS)** which are a step on the path to Autonomous Vehicles. The European automotive industry is the world leader and dominant in the market for ADAS. The technologies depend upon the analysis of video and other vehicle sensor data. **Annotations** of road traffic objects, events and scenes are critical for training and testing computer vision techniques that are the heart of modern ADAS and Navigation systems. Thus, building ADAS algorithms using machine learning techniques require annotated data sets. Human annotation is an expensive and error-prone task that has only been tackled on small scale to date. Currently no commercial tool exists that addresses the need for semi-automated annotation or that leverages the elasticity of Cloud computing in order to reduce the cost of the task. Providing this capability will establish a sustainable basis to drive forward automotive Big Data Technologies. Furthermore, the computer is set to become the central hub of a connected car and this provides the opportunity to investigate how these Big Data Technologies can be scaled to perform lightweight analysis on board, with results sent back to a Cloud Crowdsourcing platform, further reducing the complexity of the challenge faced by the Industry. Car manufacturers can then in turn cyclically update the ADAS and Mapping software on the vehicle benefiting the consumer.

<부록 21> H2020 오픈소스소프트웨어 과제 기준 예비 후보

[H2020 오픈소스소프트웨어 예비후보1]

Open-source freeware for fiber optic communication and sensing simulations
(**SIMFREE**)

http://cordis.europa.eu/project/rcn/209292_en.html

Project ID: 748767

From 2017-07-01 to 2019-06-30, ongoing project

Total cost: EUR 134 462,40 (EU contribution: EUR 134 462,40)

Coordinated in: Poland

Funded under: H2020-EU.1.3.2. - Nurturing excellence by means of cross-border and cross-sector mobility

Topic(s): MSCA-IF-2016 - Individual Fellowships

Call for proposal: H2020-MSCA-IF-2016 See other projects for this call

Funding scheme: MSCA-IF-EF-ST - Standard EF

Objective

To meet the constantly increasing traffic in optical networks due to ever growing requirements of society spectral efficiency should be increased by using, an advanced modulation formats and reduced channel spacing. This results in increased spectral power density and interferences with an adjacent channels. Furthermore, dynamic links switching in the network is foreseen. All this makes the physical layer simulations an important task in network modeling, helping to optimise the transmission range of individual optical paths. The commercial tools designed to simulate fiber optic systems, are rather expensive and there is no simpler alternative software of this type.

The ambition of this proposal is **to develop intuitive software, which can be used as a support for research projects and most importantly in education**. The basic Scilab open-source software simulator modules based on algorithms available in the literature will be developed and verified with the simulation results obtained using known commercial simulators. The aim of the project is create modules and combine techniques from digital signal processing and numerical modelling into the development of novel solutions for nonlinearity compensation in optically amplified optical transmission systems. The amplification based on advanced distributed Raman schemes can cope with the rapid development of optical telecommunication techniques and can be also applied to distributed sensing technology that enables continuous, real-time measurements along the entire length of a fibre optic cable.

This fellowship is to provide the early career, but experienced researcher Paweł Rosa with an opportunity to expand his current experimental and theoretical knowledge in the field of distributed Raman technology and develop novel freeware open-source Scilab modules for optical communications and sensing applications.

[H2020 오픈소스소프트웨어 예비후보2]

ElasTest: an elastic platform for testing complex distributed large software systems
(**ELASTEST**)

http://cordis.europa.eu/project/rcn/205931_en.html

Project ID: 731535

From 2017-01-01 to 2019-12-31, ongoing project

Total cost: EUR 5 031 187,50 (EU contribution: EUR 4 270 187)

Coordinated in: Spain

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-10-2016 - Software Technologies

Call for proposal: H2020-ICT-2016-1See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

The demand for larger and more interconnected software systems is constantly increasing, but the ability of developers to satisfy it is not evolving accordingly. The most limiting factor is software validation, which typically requires very costly and complex testing processes. This project aims at **significantly improving the efficiency and effectiveness of the testing process and, with it, the overall quality of large software systems**. For this, we propose to apply the “divide-and-conquer” principle, which is commonly used for architecting complex software, to testing by developing a novel test orchestration theory and toolbox enabling the creation of complex test suites as the composition of simple testing units. This test orchestration mechanism is complemented with a number of tools that include: (1) Capabilities for the instrumentation of the Software under Test enabling to reproduce real-world operational conditions thanks to features such as Packet Loss as a Service, Network Latency as a Service, Failure as a Service, etc. (2) Reusable testing services solving common testing problems including Browser Automation as a Service, Sensor Emulator as a Service, Monitoring as a Service, Security Check as a Service, Log Ingestion and Analysis as a Service, Cost Modeling as a Service, etc. (3) Cognitive computing and machine learning mechanisms suitable for ingesting large amounts of knowledge (e.g. specifications, logs, software engineering documents, etc.) and capable of using it for generating testing recommendations and answering natural language questions about the testing process. The ElasTest platform thus created shall be released basing on a flexible Free Open Source Software and a community of users, stakeholders and contributors shall be grown around it with the objective of transforming ElasTest into a worldwide reference in the area of large software systems testing and of guaranteeing the long term sustainability of the project generated results.

[H2020 오픈소스소프트웨어 예비후보3]

Developer-Centric Knowledge Mining from Large Open-Source Software Repositories
(**CROSSMINER**)

http://cordis.europa.eu/project/rcn/206182_en.html

Project ID: 732223

From 2017-01-01 to 2019-12-31, ongoing project

Total cost: EUR 4 519 007,50 (EU contribution: EUR 4 519 007,50)

Coordinated in: United Kingdom

Funded under: H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

Topic(s): ICT-10-2016 - Software Technologies

Call for proposal: H2020-ICT-2016-1See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

Recent reports state that the adoption of open-source software (OSS) helps, resulting in savings of about \$60 billion per year to consumers. However, the use of OSS also comes at enormous cost: **choosing among OSS projects and maintaining dependence on continuously changing software requires a large investment**. Deciding if an OSS project meets the required standards for adoption is hard, and keeping up-to-date with an evolving project is even harder. It involves **analysing code, documentation, online discussions, and issue trackers**. There is too much information to process manually and it is common that uninformed decisions have to be made with detrimental effects.

CROSSMINER remedies this by automatically extracting the required knowledge and injecting it into the IDE of the developers, at the time they need it to make their design decisions. This allows them to reduce their effort in knowledge acquisition and to increase the quality of their code. CROSSMINER uniquely combines advanced software project analyses with online monitoring in the IDE. The developer will be monitored to infer which information is timely, based on readily available knowledge stored earlier by a set of advanced offline deep analyses of related OSS projects.

To achieve this timely and ambitious goal, CROSSMINER combines six end-user partners (in the domains of IoT, multi-sector IT services, API co-evolution, software analytics, software quality assurance, and OSS forges), along with R&D partners that have a long track-record in conducting cutting-edge research on large-scale software analytics, natural language processing, reverse engineering of software components, model-driven engineering, and delivering results in the form of widely-used, sustainable and industrial-strength OSS. The development of the CROSSMINER platform is guided by an advisory board of world-class experts and the dissemination of the project will be led by The Open Group.

[H2020 오픈소스소프트웨어 예비후보4]

Statistical multi-Omics UNDERstanding of Patient Samples (**SOUND**)

http://cordis.europa.eu/project/rcn/193265_en.html

http://cordis.europa.eu/result/rcn/202073_en.html

Project ID: 633974

From 2015-09-01 to 2018-08-31, ongoing project

Total cost: EUR 3 656 977,50 (EU contribution: EUR 2 953 062,50)

Coordinated in: Germany

Funded under: H2020-EU.3.1. - SOCIETAL CHALLENGES - Health, demographic change and well-being

Topic(s): PHC-32-2014 - Advancing bioinformatics to meet biomedical and clinical needs

Call for proposal: H2020-PHC-2014-two-stage See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

Bioinformatic analysis is the biggest bottleneck in many genomic medicine projects. Our objective is **to enable researchers to dramatically increase statistically informed use of personal multi-omic data in medicine**. Soon, multiple types of omic technologies will be applied to **100,000s of patient-derived samples**, with the three-stage goal of better understanding disease biology, discovery of new interventions, and personalizing the choice of treatment options. Our interdisciplinary team of biostatisticians, bioinformaticians, software developers and physician-scientists will address the analysis bottleneck with statistically and computationally sound methods. The SOUND consortium will (i) develop widely used and excellent bioinformatic and statistical methods and open source software for common but challenging tasks, including data pre-processing, data integration, statistical inference, visual presentation, and publication-quality reporting; (ii) introduce novel approaches to ground breaking multi-omics applications in oncology and medical genetics; (iii) develop interoperable data structures and software interfaces that enable seamless combination of tools; (iv) support a collaborative international academic and industry developer community; (v) enable rapid development and execution of high-quality software; (vi) lower the barrier to entry into this transdisciplinary field by providing simple, robust, easy-to-use solutions; and (vii) develop a training programme with regular courses and comprehensive online tutorials. Our aim is to create the de facto standard toolkit used in every clinical research lab for statistically informed analysis of personal multi-omic data. SOUND will increase research and innovation opportunities by reducing barriers of entry to genomic medicine across academic, healthcare and commercial sectors by translating in a rapid and efficient manner complex and innovative analytical approaches into modular, interoperable, reusable applications.

[H2020 오픈소스소프트웨어 예비후보5]

Service Programing and Orchestration for Virtualized Software Networks (**SONATA**)

http://cordis.europa.eu/project/rcn/197338_en.html

http://cordis.europa.eu/result/rcn/191389_en.html

Project ID: 671517

From 2015-07-01 to 2017-12-31, ongoing project

Total cost: EUR 8 256 066,25 (EU contribution: EUR 6 657 721,38)

Coordinated in: Spain

Funded under: H2020-EU.2.1.1.3. - Future Internet: Software, hardware, Infrastructures, technologies and services

Topic(s): ICT-14-2014 - Advanced 5G Network Infrastructure for the Future Internet

Call for proposal: H2020-ICT-2014-2See other projects for this call

Funding scheme: IA - Innovation action

Objective

Virtualisation and software networks are a major disruptive technology for communications networks, enabling services to be deployed as software functions running directly in the network on commodity hardware. However, deploying the more complex user-facing applications and services envisioned for 5G networks presents significant technological challenges for development and deployment. SONATA addresses both issues.

For service development, SONATA provides **service patterns and description techniques for composed services**. A customised SDK is developed to boost the efficiency of developers of network functions and composed services, by integrating catalogue access, editing, debugging, and monitoring analysis tools with service packaging for shipment to an operator.

For deployment, SONATA provides **a novel service platform to manage service execution**. The platform complements the SDK with functionality to validate service packages. Moreover, it improves on existing platforms by providing a flexible and extensible orchestration framework based on a plugin architecture. Thanks to SONATA's platform service developers can provide custom algorithms to steer the orchestration of their services: for continuous placement, scaling, life-cycle management and contextualization of services. These algorithms are overseen by executives in the service platform, ensuring trust and resolving any conflict between services.

By combining rapid development and deployment in an open and flexible manner, SONATA is realising an extended DevOps model for network stakeholders.

SONATA validates its approach through novel use-case-driven pilot implementations and disseminates its results widely by releasing its key SDK and platform components as open source software, through scientific publications and standards contributions, which, together, will have a major impact on incumbent stakeholders including network operators and manufacturers and will open the market to third-party developers.

[H2020 오픈소스소프트웨어 예비후보6]

FREE and open source software tools for WATER resource management (**FREEWAT**)

http://cordis.europa.eu/project/rcn/193895_en.html

http://cordis.europa.eu/result/rcn/198380_en.html

Project ID: 642224

From 2015-04-01 to 2017-09-30, closed project

Total cost: EUR 1 583 912,50 (EU contribution: EUR 1 411 162,50)

Coordinated in: Italy

Funded under: H2020-EU.3.5.4. - Enabling the transition towards a green economy and society through eco-innovation

Topic(s): WATER-4a-2014 - Dissemination and exploitation, ICT, knowledge, gaps, research needs, etc

Call for proposal: H2020-WATER-2014-one-stageSee other projects for this call

Funding scheme: CSA - Coordination and support action

Objective

"FREEWAT aims at promoting water management and planning by simplifying the application of the Water Framework Directive and other EU water related Directives. FREEWAT will be an open source and public domain GIS integrated modelling environment for the simulation of water quantity and quality in surface water and groundwater with an integrated water management and planning module.

Specific objectives of the FREEWAT project are:

- to coordinate previous EU and national funded research to integrate existing software modules for water management in a single environment into the GIS based FREEWAT;
- to support the FREEWAT application in an innovative participatory approach gathering technical staff and relevant stakeholders (in primis policy and decision makers) in designing scenarios for the proper application of water policies.

FREEWAT will initiate a process aimed at filling the gap between EU and US on widespread-standardised ICT tools and models for management of water quantity and quality and will set a well recognisable and flagship initiative. The open source characteristics of the platform allow to consider this an initiative "ad includendum" (looking for inclusion of other entities), as further research institutions, private developers etc. may contribute to the platform development.

Through creating a common environment among water research/professionals, policy makers and implementers, FREEWAT main impact will be on enhancing science- and participatory approach and evidence-based decision making in water resource management, hence producing relevant and appropriate outcomes for policy implementation.

The Consortium is constituted by partners from various water sectors from 11 EU countries, plus Switzerland, Turkey and Ukraine. Synergies with the UNESCO HOPE initiative on free and open source software in water management greatly boost the value of the project. Large stakeholders involvement guarantees results dissemination and exploitation."

[H2020 오픈소스소프트웨어 예비후보7]

Enabling Federated Cloud Networking (**BEACON**)

http://cordis.europa.eu/project/rcn/194143_en.html

http://cordis.europa.eu/result/rcn/193056_en.html

Project ID: 644048

From 2015-02-01 to 2017-07-31, closed project

Total cost: EUR 3 570 250 (EU contribution: EUR 3 570 250)

Coordinated in: Belgium

Funded under: H2020-EU.2.1.1.3. - Future Internet: Software, hardware, Infrastructures, technologies and services

Topic(s): ICT-07-2014 - Advanced Cloud Infrastructures and Services

Call for proposal: H2020-ICT-2014-1See other projects for this call

Funding scheme: RIA - Research and Innovation action

Objective

Cloud federation enables cloud providers to collaborate and share their resources to create a large virtual pool of resources at multiple network locations. Different types of federation architectures for clouds and datacenters have been proposed and implemented (e.g. cloud bursting, cloud brokering or cloud aggregation) with different level of resource coupling and interoperation among the cloud resources, from loosely coupled, typically involving different administrative and legal domains, to tightly coupled federation, usually spanning multiple datacenters within an organization. In both situations, an effective, agile and secure federation of cloud networking resources is key to impact the deployment of federated applications.

The main goal of this project is two-fold: **research and develop techniques to federate cloud network resources**, and **to derive the integrated management cloud layer that enables an efficient and secure deployment of federated cloud applications**. Our proposal will deliver a homogeneous virtualization layer, on top of heterogeneous underlying physical networks, computing and storage infrastructures, providing enablement for automated federation of applications across different clouds and datacenters.

The project is fully committed to open source software. Cloud networking aspects will be based on OpenDaylight, a collaborative project under The Linux Foundation, and specifically we will leverage and extend the OpenDOVE project with new rich inter-cloud APIs to provision cross-site virtual networks overlays.

The new inter-cloud network capabilities will be leveraged by existing open source cloud platforms, OpenNebula and OpenStack, to deploy multi-cloud applications. In particular, different aspects of the platforms will be extended to accommodate the federated cloud networking features like multi-tenancy, federated orchestration of networking, compute and storage management or the placement and elasticity of the multi-cloud applications.

[H2020 오픈소스소프트웨어 예비후보8]

AppHub, the European Open Source Marketplace (**AppHub**)

http://cordis.europa.eu/project/rcn/194279_en.html

http://cordis.europa.eu/result/rcn/186429_en.html

Project ID: 645096

From 2015-01-01 to 2016-12-31, closed project

Total cost: EUR 849 260 (EU contribution: EUR 849 260)

Coordinated in: Germany

Funded under: H2020-EU.2.1.1.3. - Future Internet: Software, hardware, Infrastructures, technologies and services

Topic(s): ICT-07-2014 - Advanced Cloud Infrastructures and Services

Call for proposal: H2020-ICT-2014-1See other projects for this call

Funding scheme: CSA - Coordination and support action

Objective

Many software engineering and cloud computing are developed in open source in order to implicitly avoid the curse of the European science paradox -- we are good at science but poor at exploitation – but fail because they lack adequate support and strategy. Open source provides an efficient framework for cooperation and IP management and that makes it indeed a powerful enabler for collaborative innovation. It is wrong however to assume that making the source available automatically attracts contributors and grants immediate market access. Open source as a convenient process is not the same as open source as a business strategy.

The aim of this project is **to support the market outreach strategies of EU-supported open source by launching AppHub, the European open source market place**. AppHub is a service platform that will help the market to seamlessly identify, position and implement the software outcomes of these projects. The partners that will develop, run and promote AppHub over this two-year project and beyond combine unparalleled expertise in open source community management, EU research projects and a breakthrough technology in software asset management.

<부록 22> 후보 선정을 위한 의견 수렴 양식(DARPA)

○ 목적

- DARPA가 현재 진행하고 있는 사업 중 중장기 대형 과제로 추진할만한 것 추천
- 아래에서 제시한 예비 후보와 유사한 과제를 국내에서도 추진할 필요가 있는가?

○ 예비 후보 선발

- DARPA에서 진행 중인 219개 사업 대상 (2017.8.20. 기준)
- SW 관련 분야 사업 중심으로 검토 (분류 예: **AI**, Algorithms, **Analytics**, Automation, Autonomy, **Data**, Electronics, Fundamentals, Imagery, Language, Manufacturing, Math, Microchips, Privacy, Processing, **Programing**, Quantum, **Resilience**, Robotics, **Systems**, Trust, Unmanned, Visualization)
 - ※ 위에 제시된 분류 모두 SW 관련 분야임. 특히 굵은 글씨는 SW 관련성이 높은 분야

○ 예비 후보 (단, 연구개발 목표 등 세부 내용은 선정 후 별도 기획 필요)

- [예비후보1] Active Interpretation of Disparate Alternatives (AIDA)
- [예비후보2] Explainable Artificial Intelligence (XAI)
- [예비후보3] Data-Driven Discovery of Models (D3M)
- [예비후보4] Synergistic Discovery and Design (SD2)
- [예비후보5] Probabilistic Programming for Advancing Machine Learning (PPAML)
- [예비후보6] Mining and Understanding Software Enclaves (MUSE)
- [예비후보7] SAFEWARE
- [예비후보8] Building Resource Adaptive Software Systems (BRASS)
- [예비후보9] Hierarchical Identify Verify Exploit (HIVE)

○ 선정 기준

- 소프트웨어 연구개발 과제로서 적절한가? (과제 내 소프트웨어의 비중 등)
- 기술 격차를 획기적 해소하는데 기여할 수 있는가?
- 중장기 대형 과제로 추진할 필요가 있는가? (5년 이상 연 50억 원 이상 투자 필요)

○ 검토의견

○ 적극 추천 후보 (3개 이내 권장):

○ 배제 희망 후보:

○ 기타 자유로운 의견:

<부록 23> 절대비교 후보 제안요청서

(지능정보-총괄/1세부) 자율지능 디지털 동반자 프레임워크 및 응용 연구개발

(KETI 정혜동, 2017.4.1.~2020.12.31. 2017년도 예산 50.0억원)

1. 연구목표

□ **총괄 목표** : 사용자를 이해하고 적절한 도움을 주는 동반자 관점에서, 다양한 입력*을 수집하여 학습하는 적응형 기계학습*(Adaptive Machine Learning) 기반 자율지능 디지털 동반자 기술 개발

* 입력 예시: 정형/비정형데이터(음성, 영상, 텍스트, 공간정보, 장치정보, 위치정보, 센서 데이터 등)

** 적응형 기계학습: 최초 학습된 모델을 활용하면서 추가적으로 수집, 인식되는 데이터(예: 개별 사용자 특성, 습관, 어투, 용모 등과 사용자의 주요 동선, 관계 등 부수적 정보)를 지속적으로 축적하면서 점진적이고 신속하게 학습하고 적응하여 최초 모델의 품질을 개량하는 방법

○ **적응형 기계학습 기반 자율지능 디지털 동반자 프레임워크 연구 개발**

- 사람과 다양한 방식으로 교류하며, 상대의 의도 파악 및 감정까지 인지하여 도우미 역할을 할 수 있는 디지털 동반자 기술 연구개발
- 다양한 입력 방법을 수용하여 학습하고, 지속적으로 성능을 개선하는 적응형 기계학습 방법 연구개발
- 자율지능 동반자 프레임워크 기반 응용 서비스 연구 개발

□ **1세부 최종 목표** : 스스로 상황을 판단하고, 대응방법을 결정할 수 있는 **자율지능 디지털 동반자 프레임워크**와 이를 활용한 **응용/서비스** 개발

○ 단순한 미션 중심의 수동 에이전트 개념을 넘어, 지속적인 사용자 이해와 도움을 주는 디지털 동반자를 생성·실행·갱신·소멸할 수 있는 디지털 동반자* 프레임워크 기술 확보

* 디지털 동반자 주요 성질: 자율성, 사회성, 반응성, 능동성, 시간 연속성, 상황 적응성 등

○ 디지털 동반자의 효과가 큰 응용 도메인을 선정하여 각 세부과제의 결과*를 적용한 응용/서비스 개발

* 1세부: 디지털 동반자 프레임워크, 2세부: 적응형 기계학습, 3세부: 지능형 인터랙션, 4세부: 동반자 감성지능 기술

○ 사용자와 공감하고 실생활에 도움을 주는 디지털 동반자 대국민 시연 및 시범 서비스

* 응용도메인 예시: 운전 중 주변 시정각 정보를 이해하여 운전자를 도와주는 주행 상황판단 동반자, 지속적인 사용자 이해기반의 챗봇, 고객 맞춤형 자산관리 컨설팅 등

□ 개발목표

핵심 기술/제품 성능지표		단위	달성목표	국내최고수준	세계최고수준 (보유국, 기업/기관명)
1	동반자의 멀티모달 자율지능 정도 (동반자 튜링 테스트 ¹⁾)	%	33	N/A	N/A
2	디지털 동반자 사용자 만족도 (리커트 척도 ²⁾)	점수	3.5 이상	N/A	N/A
3	동반자 공개 API 제공 여부 (주요 공통 기능)	지원여부	지원	-	-

1) 튜링테스트 개념을 활용하여 30명 이상의 사용자를 대상으로 인간과 디지털 동반자를 언어, 감성, 인터랙션 등의 측면에서 구분하지 못하는 사람이 모수의 3분의 1 이상이면 통과 * 언어지능의 경우 세계 최고 수준: 일반대화 33%, 13세, 비영어권 기준(영국, 레딩대, 30명 모수 사용)

2) 디지털 동반자 응용을 사용해본 사용자에게 기본 5단계*로 설문조사하여 평균 집계*

* 5단계: 1-매우불만족, 2-불만족, 3-보통, 4-만족, 5-매우만족

** 평가 항목은 디지털 동반자 주요 성질(자율성, 사회성, 반응성, 능동성, 시간 연속성, 상황 적응성 등)로 구성

□ 연구내용

○ **‘자율지능 디지털 동반자’ 통합 시스템** 구성 기술 (총괄)

- 디지털 동반자를 형성하는 세부과제 기술의 통합 아키텍처와 인터페이스 정의
- 다양한 지능이 유기적으로 결합된 통합 시스템 개발

○ **디지털 동반자 프레임워크** 기술

- 지속성 확보를 위한 사용자와 디지털 동반자 통합 관리 프레임워크 기술
- 적응형 기계학습 기반의 자율지능 디지털 동반자 모델링 기술
- 주변에서 발생하는 다양한 상황 인지 및 평가 기술
- 사용자 상황에 따른 대응 결정을 위한 사전 플래닝 기술
- 디지털 동반자 생성 → 실행 → 갱신(적응) → 소멸 관리 기술
- 동반자 모니터링 기술
- 클라우드-동반자 간 업무분배를 통한 디바이스 동반자 경량화 설계 기술

○ 디지털 동반자 응용/서비스 개발을 위한 **공개 API 제공** 기술

- 재사용 가능한 디지털 동반자의 공개 API 제공 기술
- 공개 API 활용의 확산 및 지속적이고 점진적인 개선 체계 확보

○ 디지털 동반자 통합시스템의 **웰니스 응용/서비스** 등을 통한 기술 증명

- 세부과제들의 기술을 융합·활용하여 사용자 웰니스 분야 컴패니언 등의 시제품 개발 및 응용 서비스 구현
- 선정된 응용을 위한 디지털 동반자 플래닝, 추론, 상황판단, 클라이언트SW 기술
- 전체 시스템 또는 요소기술의 공개 경진대회 참여 등 기술의 성능 증명 방법 제시

○ 자율지능 디지털 동반자를 위한 학습 및 시험 데이터 확보

- 자율지능동반자의 적응형 기계학습 기술 개발을 위한 데이터의 종류 정의
- 공개데이터, 기관과 기업 보유 데이터 중 공유 가능한 데이터 조사 및 확보
- 신규 데이터 생성

※ 선행 정부 R&D 결과물과 오픈 프레임워크 활용을 포함한 연차별 기술 확보 계획 제시

※ 연차별 핵심 결과물 도출 계획 제시

<부록 24> 최종후보 선정을 위한 응답 양식 및 후보 설명(1차 검토 기준)

☐ 목적 및 후보 선정

○ 목적

- “아래에서 제시한 후보와 유사한 과제를 국내에서도 기획해서 추진할 필요가 있는가?”
- 단, 연구개발 목표 등 세부 내용은 선정 후 별도 기획 필요

○ 후보 선정

- 219개 DARPA 진행 사업에서 선발한 9개 예비 후보 중 4개 후보 선정
- 35개 IARPA 진행 사업에서 선발한 4개 예비 후보 중 1개 후보 선정

☐ 후보

○ DARPA

- [후보1] Active Interpretation of Disparate Alternatives (AIDA)
- [후보2] Synergistic Discovery and Design (SD2)
- [후보3] Data-Driven Discovery of Models (D3M)
- [후보4] Probabilistic Programming for Advancing Machine Learning (PPAML)

○ IARPA

- [후보5] Deep Intermodal Video Analytics (DIVA)

<후보별 수행기간 및 예산>

구분	기간 (개월)	예산 (백만 달러)
AIDA	'18~ (54)	('17) 5.5 ('18) 17.3
SD2	'17.8.23~	('17) 13.0 ('18) 21.0
D3M	'17.5~'21.4 (48)	('17) 20.247 ('18) 26.84
PPAML	'13.10~'17.7(46)	총 44.328
DIVA	'17.6.30 선정 예정 (48)	-

☐ 선정 기준

- 기술 선도국과의 기술 격차를 획기적으로 해소할 수 있다 (기술 격차 해소 가능성)
- 우리나라 연구진이 개발에 성공할 것이다 (기술개발 성공 가능성)

□ 전문가 설문 응답지

본 설문은 ‘중장기 대형 SW R&D과제 발굴’을 위한 것입니다. 이때 중장기 대형 SW R&D과제라고 함은 5년 이상 연 50억원 이상의 정부 재원이 투자되는 과제를 의미합니다.

설문은 평가항목간 상대적 중요도에 관한 것과 평가항목별 과제 추진의 필요성에 관한 것으로 구성되어 있습니다.

응답의 일관성이 낮은 경우 피드백 과정을 거치되 되오니 전문가의 관점에서 공정하고, 신중하게 응답하여 주시기 바랍니다.

각 후보들에 대한 설명은 별도 자료를 참고하여 주시기 바랍니다.

1. 중장기 대형 SW R&D과제의 기획 대상을 결정하는데 아래의 평가항목이 각각 얼마나 중요합니까? 100점 만점으로 배분해 주세요.

- (1) 선진국과의 기술 격차를 개선 또는 역전하는데 기여해야 한다..... ()점
 (2) 우리나라 연구진이 과제를 수행할 역량을 갖추고 있어야 한다..... ()점

<참고: 우리나라의 SW 분야 기술수준 (출처: IITP, ICT 기술수준조사)>

분류		기술격차(년)		기술수준(%)	
대분류	중분류	2016	2015	2016	2015
기반SW·컴퓨팅	전체	1.8	2.0	71.7	74.1
	컴퓨팅 시스템	2.1	2.0	75.5	76.6
	High Performance컴퓨팅	-	2.8	-	66.3
	인공지능·인지컴퓨팅	2.2	2.4	73.9	70.5
	클라우드 컴퓨팅	1.6	1.8	72.4	77.0
	모바일 컴퓨팅	-	1.7	-	77.9
	빅데이터 플랫폼	1.7	1.6	78.2	76.3
SW	전체	1.9	1.9	76.7	76.8
	시스템 SW (OS, DBMS 등)	2.1	2.3	74.6	72.7
	미들웨어	1.9	2.0	76.9	76.3
	응용 SW	1.8	1.3	78.7	81.5

2. 아래의 후보들에 대해 중장기 대형 SW R&D과제의 기획 대상으로 선정하는데 적합한가를 각 평가항목별로 쌍대 비교를 해 주세요. 각 후보에 대한 상세 설명은 별도자료를 참고하시기 바랍니다.

[후보1] Active Interpretation of Disparate Alternatives (AIDA): 멀티모달 학습·통합 추론

[후보2] Synergistic Discovery and Design (SD2): 데이터 기반 과학 발견 툴

[후보3] Data-Driven Discovery of Models (D3M): 도메인 전문가 대상 사용 편의 분석 도구

[후보4] Probabilistic Programming for Advancing Machine Learning (PPAML)

[후보5] Deep Intermodal Video Analytics (DIVA): 멀티 카메라 동영상 속 동작 인식 및 이해

(1) 두 후보 중 기술 선도국과의 기술 격차를 해소하거나 역전하는데 기여할 가능성이 어느 쪽이 얼마나 높습니까? 1.0~3.0배 사이의 값으로 응답해 주세요.

더 높다	3.0 배	2.5 배	2.0 배	1.8 배	1.6 배	1.4 배	1.2 배	1.1 배	동일	1.1 배	1.2 배	1.4 배	1.6 배	1.8 배	2.0 배	2.5 배	3.0 배	더 높다
후보1																		후보2
후보2																		후보3
후보3																		후보4
후보4																		후보5
후보5																		후보1

(2) 두 후보 중 우리나라 연구진이 과제를 수행할 역량을 갖추고 있을 가능성이 어느 쪽이 얼마나 높습니까? 1.0~3.0 사이의 값으로 응답해 주세요.

더 높다	3.0 배	2.5 배	2.0 배	1.8 배	1.6 배	1.4 배	1.2 배	1.1 배	동일	1.1 배	1.2 배	1.4 배	1.6 배	1.8 배	2.0 배	2.5 배	3.0 배	더 높다
후보1																		후보2
후보2																		후보3
후보3																		후보4
후보4																		후보5
후보5																		후보1

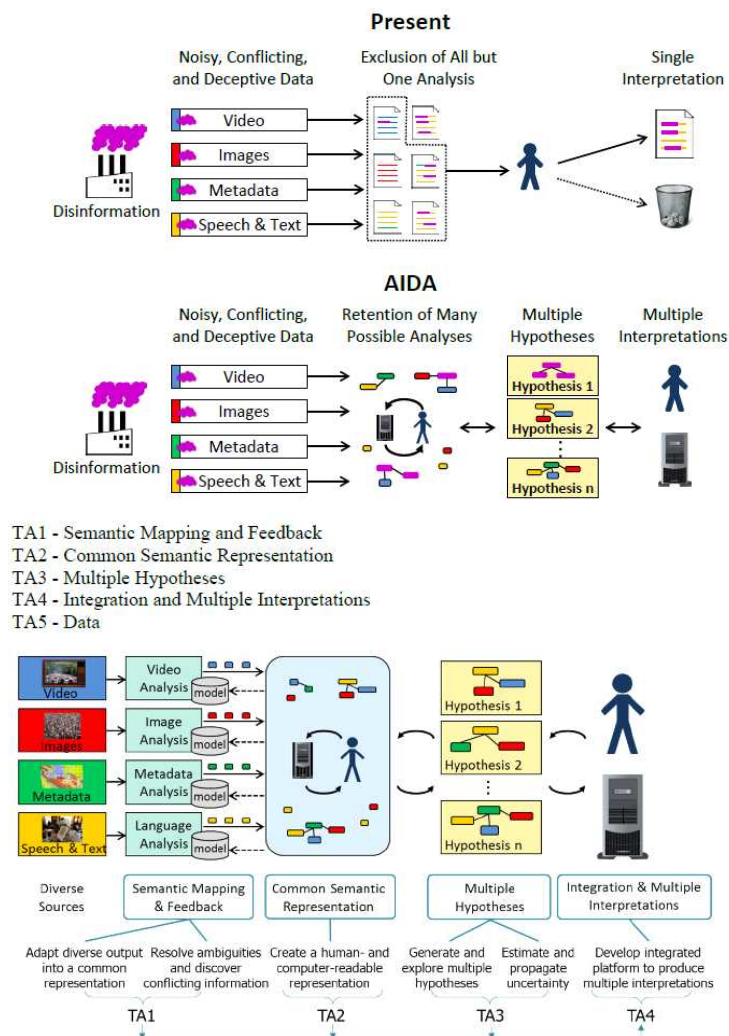
3. 평가항목, 후보 비교 등에 대해 자유로운 의견을 부탁드립니다. (선택 사항)

[제1차 후보1] Active Interpretation of Disparate Alternatives (AIDA)

PM: DR. BOYAN ONYSHKEVYCH ※ '17.6.6 제안서 접수 마감, 54개월 수행, ('17) 5.5 ('18) 17.3백만달러

다양한 비정형화된 데이터를 동시에 표현하고 분석하여 사건, 상황, 트렌드 등에 대한 복수 가설을 설정하여 의사결정자를 지원하는 엔진 개발

The goal of Active Interpretation of Disparate Alternatives (AIDA) is to develop a **multihypothesis semantic engine** that generates explicit alternative interpretations of events, situations, and trends from a variety of unstructured sources, for use in noisy, conflicting, and potentially deceptive information environments.



TA1-TA3 복수 경쟁 수행 예상, TA4-TA5 단일 수행 예상

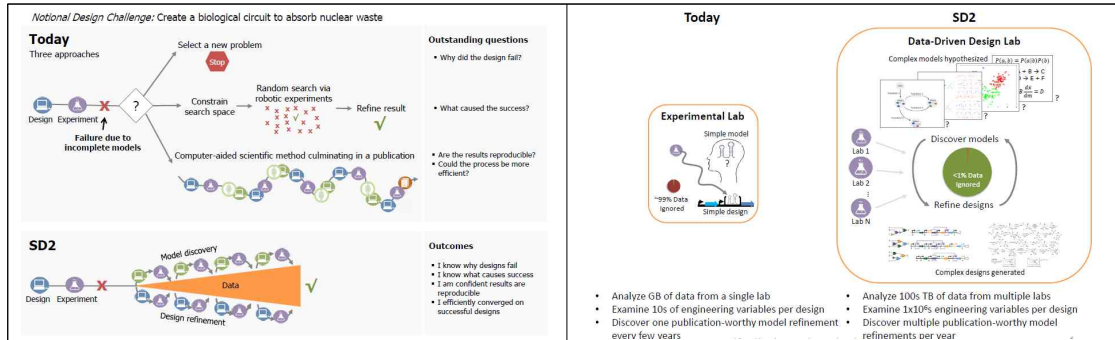
○ 국내 주요 관련 연구

- 이종 빅데이터 통합 분석을 위한 메타러닝 모델 및 응용 연구 (동국대 김준태, '17.11.1.~'20.12.31., 4,000만원)
- 딥러닝 기법을 이용한 이종 빅데이터 통합 처리 모델 개발 (중앙대 임창원, '17.11.1.~'20.12.31., 3,500만원)

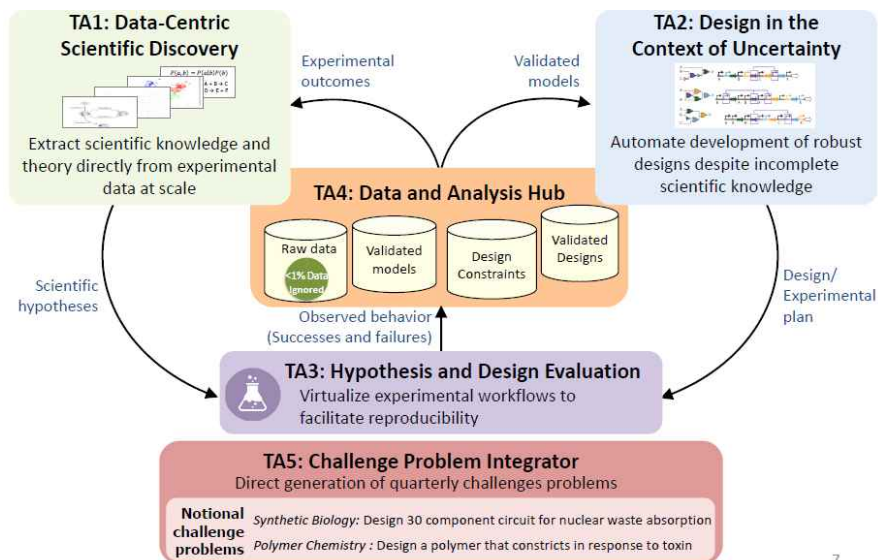
[제1차 후보2] Synergistic Discovery and Design (SD2)

PM: Dr. Jennifer Roberts ※ '17.8.23~ , ('17) 13.0 ('18) 21.0 백만달러

해당 분야에 아직 검증된 과학적 설명 모형이 존재하지 않음에도 불구하고 대용량 데이터를 이용해서 견고하게 실험 설계를 할 수 있는 방법을 고안하는 것을 목표로 하고 있고, 이를 통해 과학 발견을 가속화하고자 함 (5가지 TA로 구성됨)



The Synergistic Discovery and Design (SD2) program aims to develop data-driven methods to accelerate scientific discovery and robust design in domains that lack complete models. Engineers regularly use high-fidelity simulations to create robust designs in complex domains such as aeronautics, automobiles, and integrated circuits. In contrast, robust design remains elusive in domains such as synthetic biology, neuro-computation, and polymer chemistry due to the lack of high-fidelity models. SD2 seeks to develop tools to enable robust design despite the lack of complete scientific models.



○ 국내 주요 관련 연구

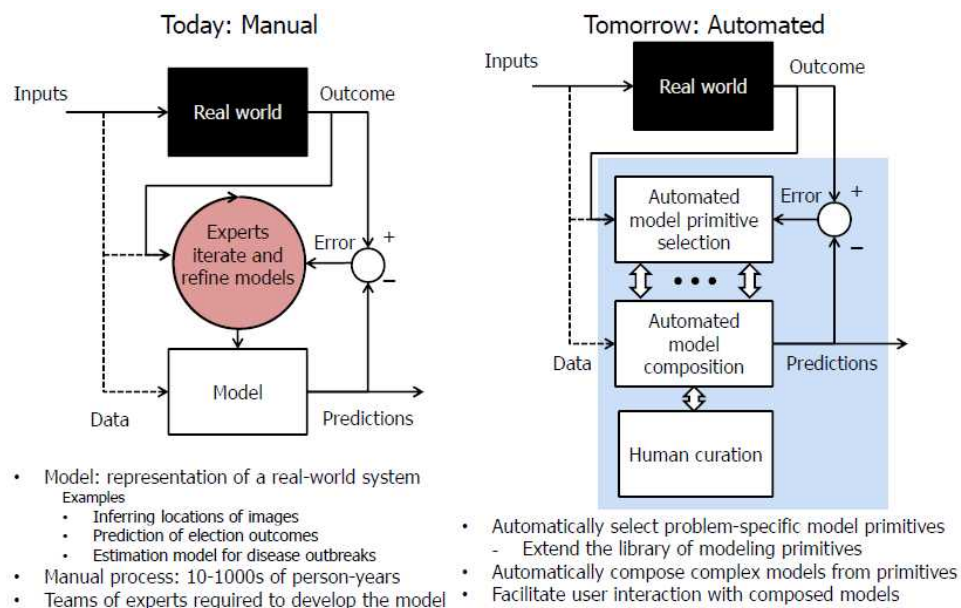
- 과학기술 빅데이터 공유 융합체제 구축 (KISTI 이상환, '14.1.1~, 53억5,800만원/41억3,400만원/41억7,200만원/35억4,800만원, 기관고유사업)
- 신소재 예측과 개발을 위한 수리과학 기반 데이터분석 기술연구개발 (국가수리과학연구소 현윤경, '17.7.1~'22.6.30, 2억2,500만원)

[제1차 후보3] Data-Driven Discovery of Models (D3M)

PM: MR. WADE SHEN ※'16.6.15 공지, 17.5~'21.4 (48) ('17) 20.247 ('18) 26.84 백만달러

데이터 과학 지식이 부족한 도메인 전문가들이 실증 연구를 할 수 있는 자동화된 모델 발견 시스템 개발

The Data-Driven Discovery of Models (D3M) program aims to **develop automated model discovery systems that enable users with subject matter expertise but no data science background to create empirical models of real, complex processes**. This capability will enable subject matter experts to create empirical models without the need for data scientists, and will increase the productivity of expert data scientists via automation. The D3M automated model discovery process, depicted in the figure, will be enabled by three key technologies to be developed in the course of the program:



- **TA1: A library of selectable primitives.** A discoverable archive of data modeling primitives will be developed to serve as the basic building blocks for complex modeling pipelines.
- **TA2: Automated composition of complex models.** Techniques will be developed for automatically selecting model primitives and for composing selected primitives into complex modeling pipelines based on user-specified data and outcome(s) of interest.
- **TA3: Human-model interaction** that enables curation of models by subject matter experts. A method and interface will be developed to facilitate human-model interaction that enables formal definition of modeling problems and curation of automatically constructed models by users who are not data scientists.

○ 국내 주요 관련 연구

- 빅 데이터 분석모델 개발과 응용 (국가수리과학연구소 현윤경, '16.1.1~, 5억6,500만원/6억7,500만원)

[제1차 후보4] Probabilistic Programming for Advancing Machine Learning (PPAML)

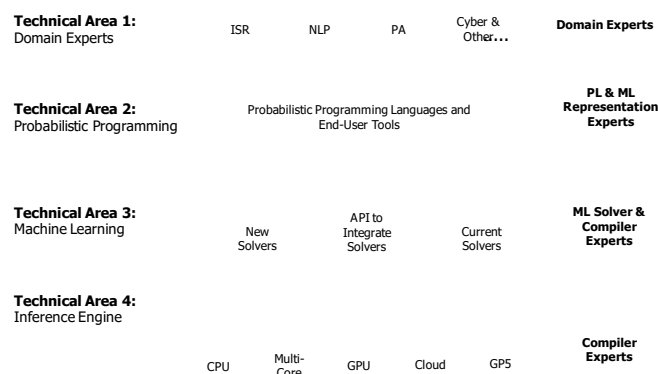
PM: Kathleen Fisher / Dr. Jennifer Roberts ※ '13.10.~'17.7 (46개월) 총 44.328 백만달러

프로그래밍 스킬이 낮은 사용자도 쉽게 사용할 수 있는 확률적 프로그래밍 패러다임 기반의 기계학습 툴 개발 (프로그래밍 언어, 컴파일러, 재사용 가능 모델 라이브러리 포함)

<i>Traditional Programming</i>	<i>Probabilistic Programming</i>	
• Application	• Model	Code models capture how the data was generated using random variables to represent uncertainty
• Code Libraries	• Model Libraries	Libraries contain common model components: Markov chains, deep belief networks, etc.
• Programming Language	• Probabilistic Programming Language	PPL provides probabilistic primitives & traditional PL constructs so users can express model, queries, and data
• Compiler	• Inference Engine	Inference engine analyzes probabilistic program and chooses appropriate solver(s) for available hardware
• Hardware	• Hardware	Hardware can include multi-core, GPU, cloud-based resources, GraphLab, UPSIDE/Analog Logic results, etc.
High-level programming languages facilitate building complex systems		
Probabilistic programming languages facilitate building rich ML applications		

The Probabilistic Programming for Advancing Machine Learning (PPAML) program aims to address these challenges. Probabilistic programming is a new programming paradigm for managing uncertain information. Using probabilistic programming languages, PPAML seeks to greatly increase the number of people who can successfully build machine learning applications and make machine learning experts radically more effective. PPAML has five specific tactical objectives:

- Shorten machine learning model code to make models faster to write and easier to understand
- Reduce development time and cost to encourage experimentation
- Facilitate the construction of more sophisticated models that incorporate rich domain knowledge and separate queries from underlying code
- Reduce the level of expertise necessary to build machine learning applications
- Support the construction of integrated models across a wide variety of domains and tool types



주) Surveillance and Reconnaissance (ISR), Natural Language Processing (NLP), Predictive Analytics (PA)

○ 관련 제품: Automatic Statistician (<https://www.automaticstatistician.com/>)

○ 국내 주요 관련 연구: 없음

[제1차 후보5] Deep Intermodal Video Analytics (DIVA)

PM: Terry Adams ※ '17.6.30 선정 예정, '17.7~'21.6 예상 (48개월)

복수의 스트리밍 카메라 환경에서의 자동 객체 및 행위 인식 방법 개발

DIVA program seeks to develop **robust automatic activity detection** for a **multi-camera streaming video environment**. Activities will be enriched by **person and object detection**. DIVA will address activity detection for both **forensic applications** and for **real-time alerting**.

- Two main goals are evaluated:
 - (a) **Activity detection** - The DIVA program intends to develop robust automated activity detection for a multi-camera streaming video environment.

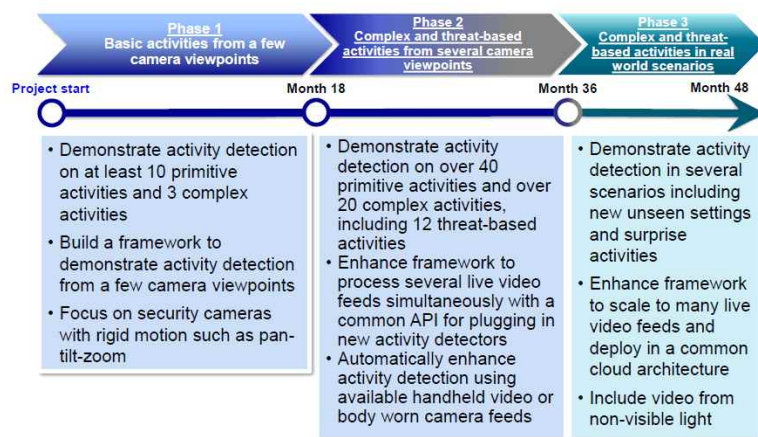


- (b) **Person/object detection** - DIVA activities will be enriched by person and object detection, as well as recognition at multiple levels of granularity.



DIVA is anticipated to be a three-phase program. The program will focus on three major thrusts throughout all phases:

- Detection of **primitive activities** occurring in ground-based video collection
- Detection of **complex activities**, including pre-specified or newly defined activities
- Person and object detection and recognition across multiple overlapping and non-overlapping camera viewpoints



○ 국내 주요 관련 연구

- (답류1세부) 실시간 대규모 영상 데이터 이해·예측을 위한 고성능 비주얼 디스커버리 플랫폼 개발 (ETRI 박경/박종열, '14.4.1~'24.2.29, 38.0억원/38.0억원/35.5억원/35.5억원)
- (답류2세부) 대규모 실시간 영상 이해 기반의 시각 지능 플랫폼 개발 (광주과학기술원 전문구, '14.4.1~'24.2.29, 20.0억원/20.0억원/18.7억원/18.7억원) *다수 카메라 등장 동일 객체 전역적 추적
- 비중첩 다중카메라 환경에서 답러닝 기반 보행자 동선 추적 요소 기술 연구 (전북대 박동선, '16.11.1~'19.10.31, 5,000만원/4,167만원)

<부록 25> DARPA AIDA 시나리오 예시: 러시아 크림반도 침공

What is Russia doing in Eastern Ukraine?

Single interpretation:

Russia is only carrying out a regular military exercise inside its own borders

Evidence for single interpretation:

Disinformation

"...naval exercises Moscow says were pre-scheduled and not timed to coincide with the current NATO drills in the basin." –Russia Today



© RT

Exclusion of conflicting information due to cognitive bias

U.S. defense officials said they saw nothing to suggest that Moscow was preparing any offensive moves. –Wall Street Journal



© Wall Street Journal

Alternate interpretation:

Russian army is active also in Eastern Ukraine

Evidence for alternate interpretation:

Russian T72 tanks

была опубликована запись с видео, на котором зафиксирован российский танк Т-72Б3 вместе с танками боевиков на территории аэропорта Луганска во время боевых действий.



© Andriy Nurzhynskyy

Uniforms with no insignia

Armed men in military uniform, believed to be Russian soldiers, block the Ukrainian navy base in Novoozerniy village, Ukraine, 03 March 2014.



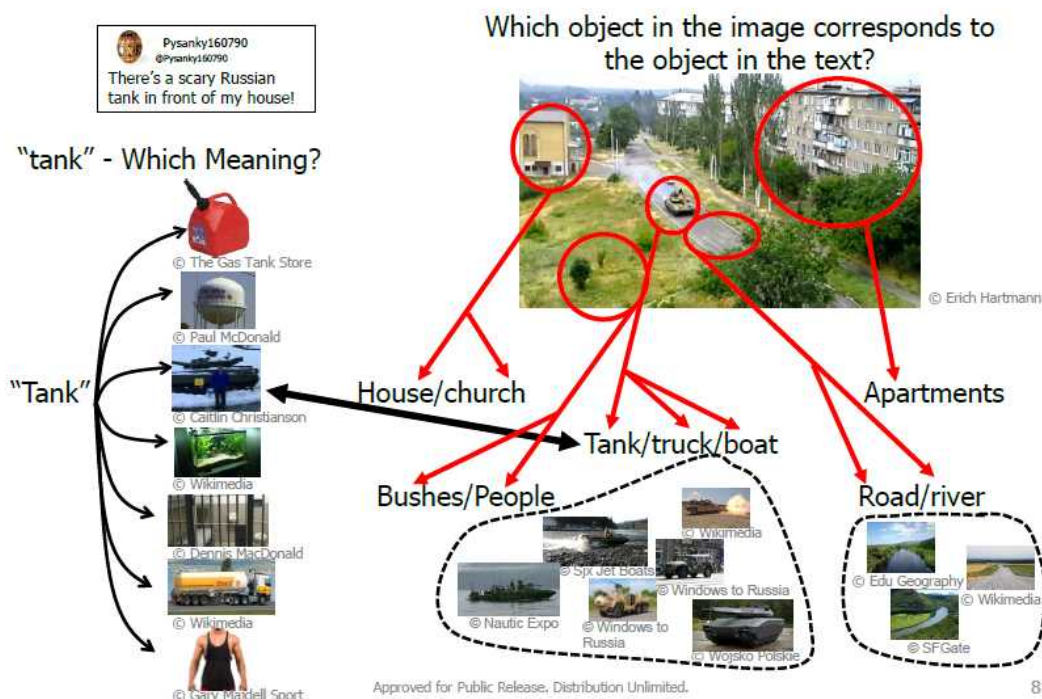
© AP

Consequences of ignoring conflicting information:


U.S. unprepared for Russian annexation of Crimea and covert support of Luhansk and Donetsk "rebellion"

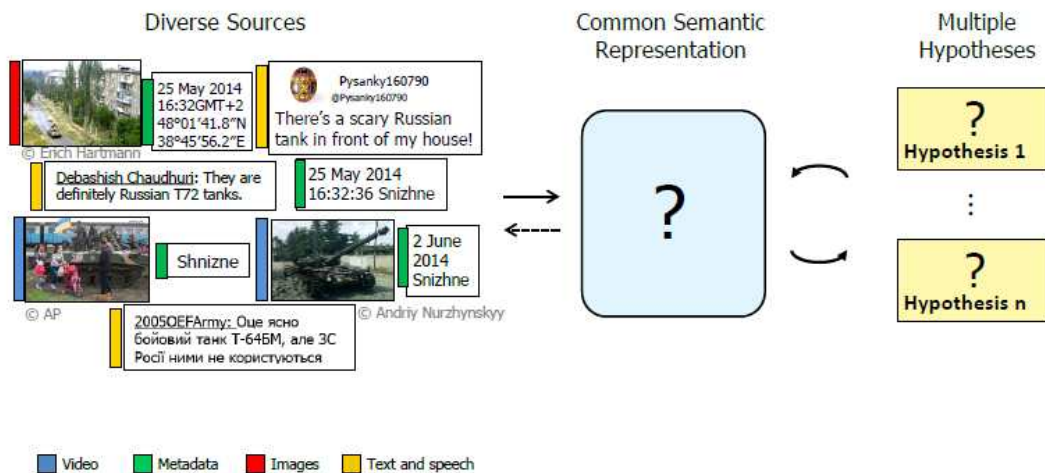


Cross-medium Disambiguation

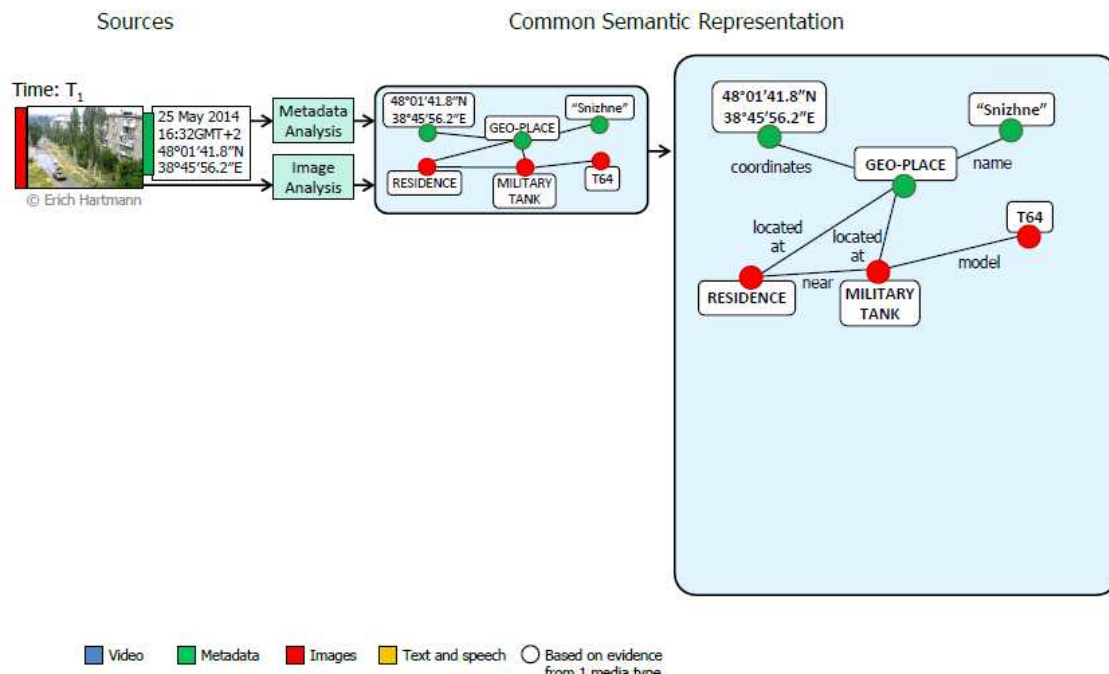


The next few slides will show an example of how AIDA will work.
The example starts with a human submitting a query to the system.

 : What is Russia doing in Eastern Ukraine?

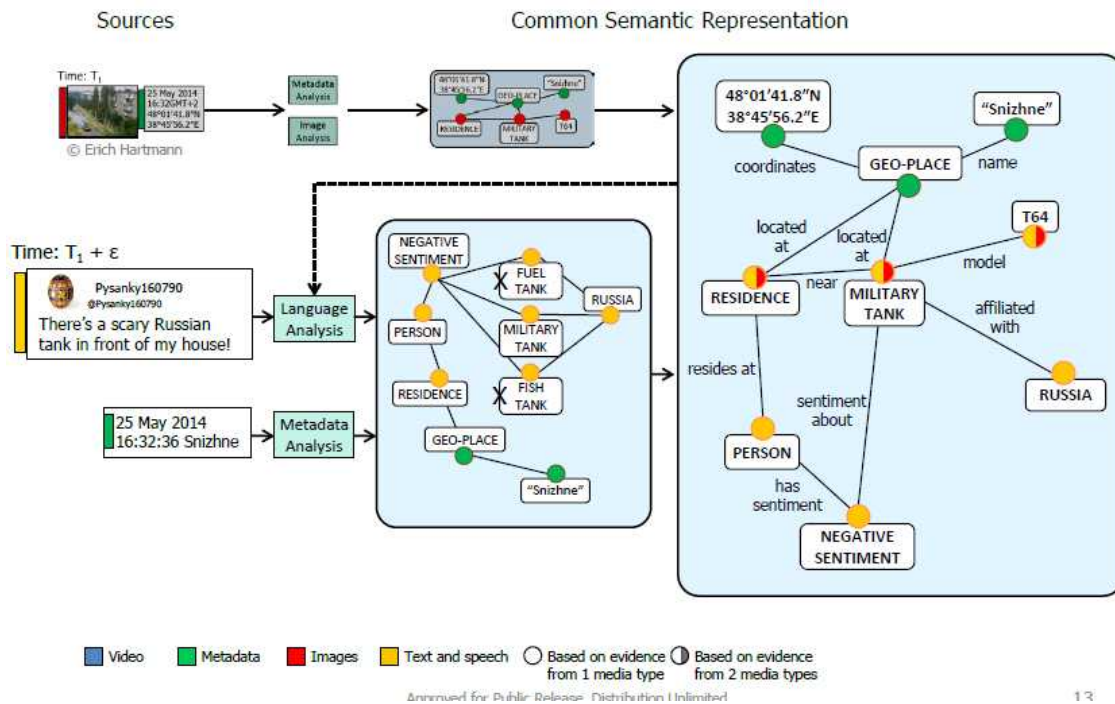


Semantic Mapping into a Common Representation





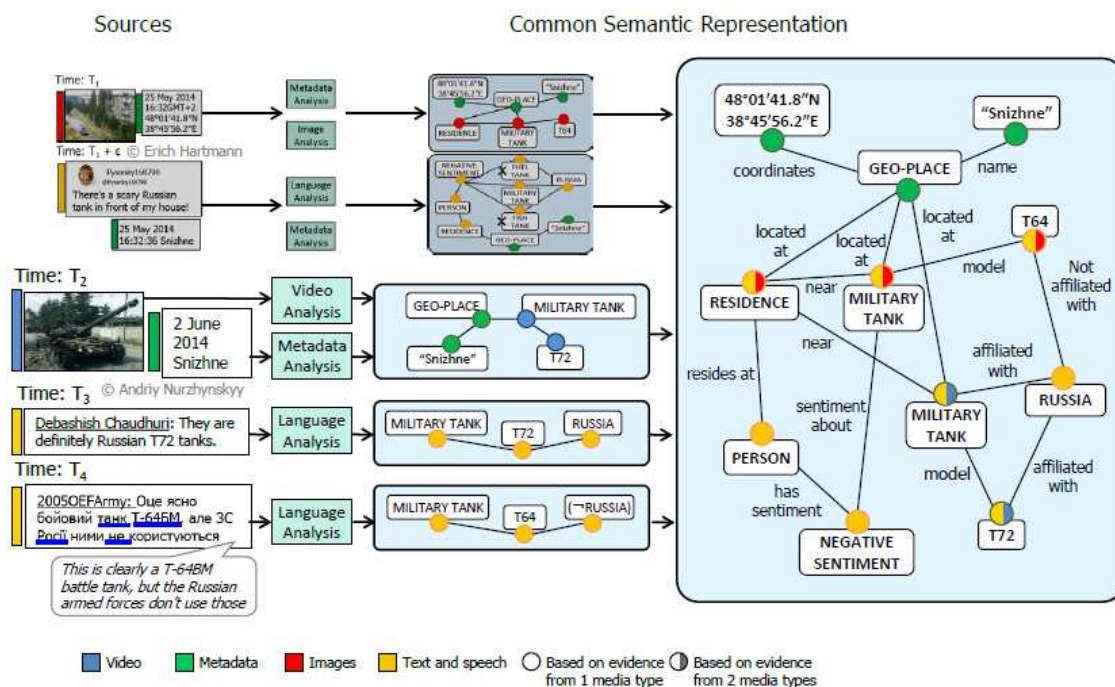
Semantic Feedback for Ambiguity Resolution



13

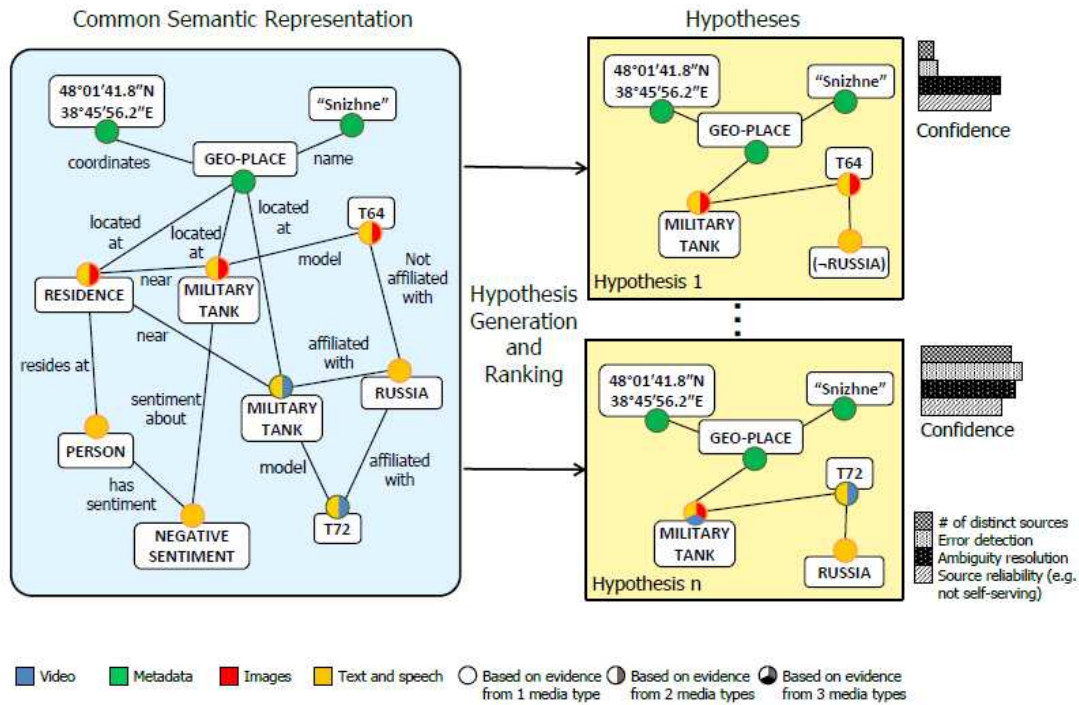


Discovery and Mapping of Conflicting Information





Generation of Hypotheses with Confidence Scores



주 의

1. 이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.



[소프트웨어정책연구소]에 의해 작성된 [SPRI 보고서]는 공공저작물 자유이용허락 표시기준 제 4유형(출처표시-상업적이용금지-변경금지)에 따라 이용할 수 있습니다.
(출처를 밝히면 자유로운 이용이 가능하지만, 영리목적으로 이용할 수 없고, 변경 없이 그대로 이용해야 합니다.)