

특허관점에서 본 보건의료부문 빅데이터 및 적용사례

2013. 10. 02

특허법인 이상
이학규 부장
esang@espat.co.kr

발표내용

I. 빅데이터 정의

II. 보건의료부문 빅데이터 출원동향

III. IBM 보건의료부문 빅데이터 특허출원 분석

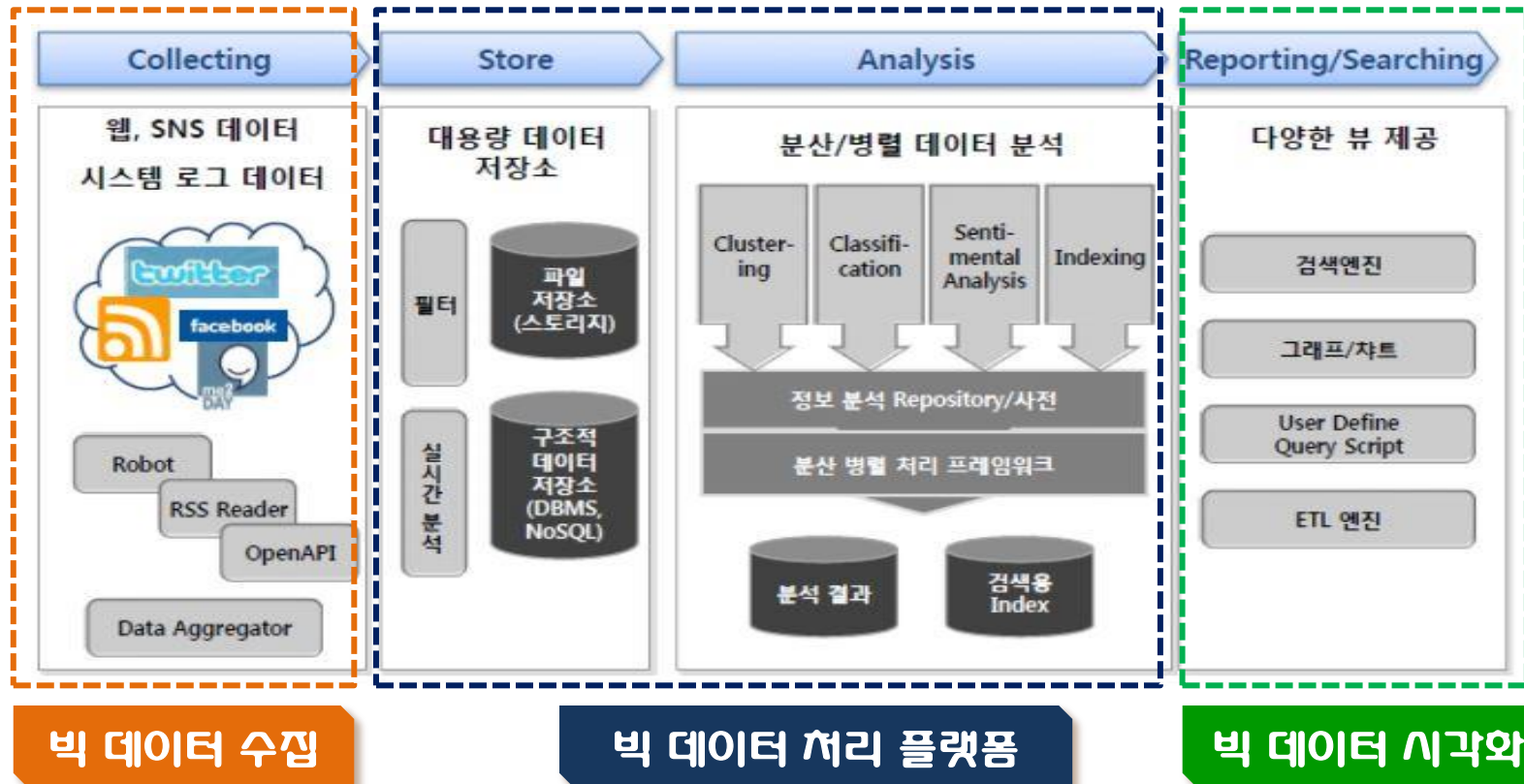
IV. 보건의료부문 빅데이터 적용사례

I. 빅 데이터(Big Data) 개념 – 정의 및 구성요소 (1/2)

	빅 데이터 정의
매킨지 (McKinsey)	전통적인 데이터베이스 소프트웨어로는 수집, 저장, 관리, 분석이 어려울 정도의 큰 규모의 데이터
IDC	다양한 데이터로 구성된 방대한 볼륨의 데이터로부터 고속 캡처, 데이터 탐색 및 분석을 통해 경제적으로 필요한 가치를 추출 할 수 있도록 디자인된 차세대 기술과 아키텍처
가트너 (Gartner)	대량의 데이터가 실시간으로 끊임없이 다양한 형태 로 들어오는 것

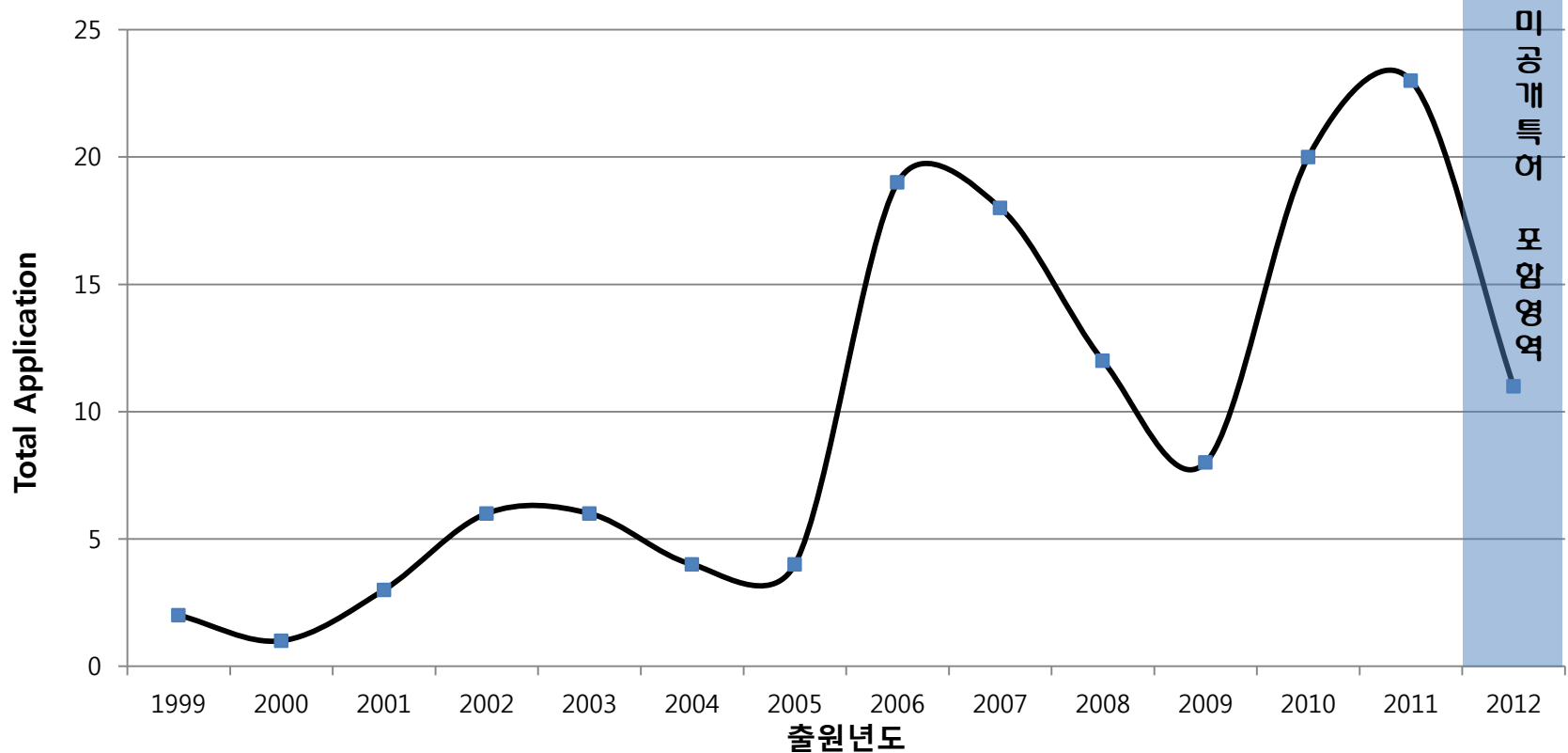
	구성	기존 데이터	빅 데이터 (Big Data)
기본 구성 요소	데이터 규모 (Volume)	제한된 용량	무한용량
	데이터 다양성 (Variety)	정형적 데이터 형태가 많음 (정형적 > 비정형적)	비정형적 데이터 형태가 많음 (정형적 < 비정형적)
	데이터 속도 (Velocity)	Batch형	실시간 + Batch
데이터 분석	데이터 분석 (Data Analysis)	통계분석 방법 사용 (사물 서술형 데이터)	통계분석 방법 사용 (사물 서술형 + 관계형 데이터)
	데이터 시각화 (Data Visualization)	데이터 프레젠테이션	인사이트(Insight)를 나타내는 Infor-graph 사용

빅 데이터(Big Data) 개념 – 정의 및 구성요소 (2/2)



- 빅 데이터는 많게는 6단계(수집, 저장, 관리, 처리, 분석 및 시각화) 혹은 적게는 4단계(수집, 저장, 분석, 시각화)로 표현할 수 있으나, 빅 데이터가 처리되는 **플랫폼 중심으로 보면 3단계로 나눌수 있음**

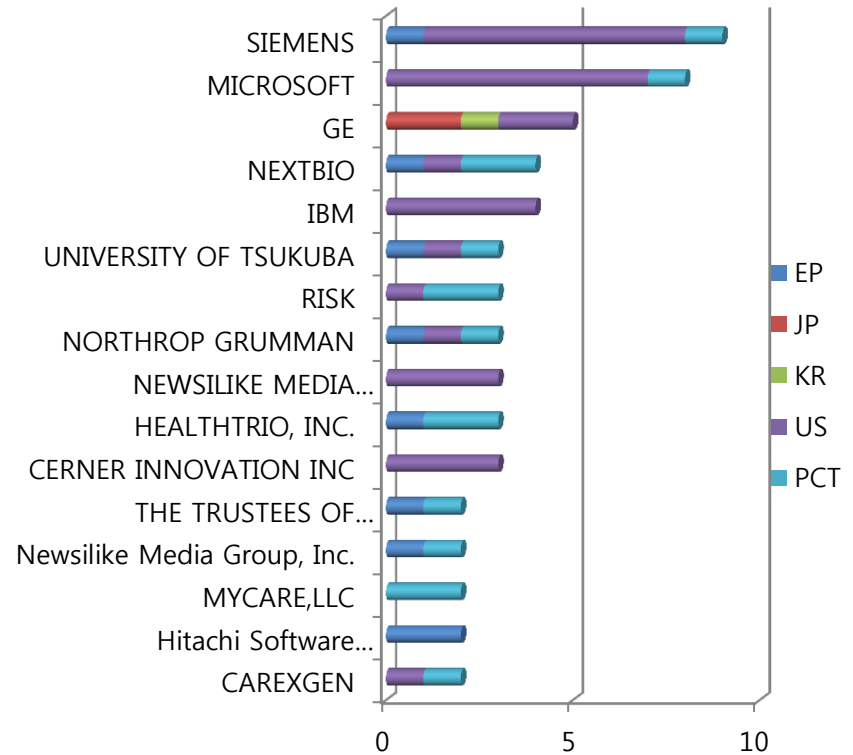
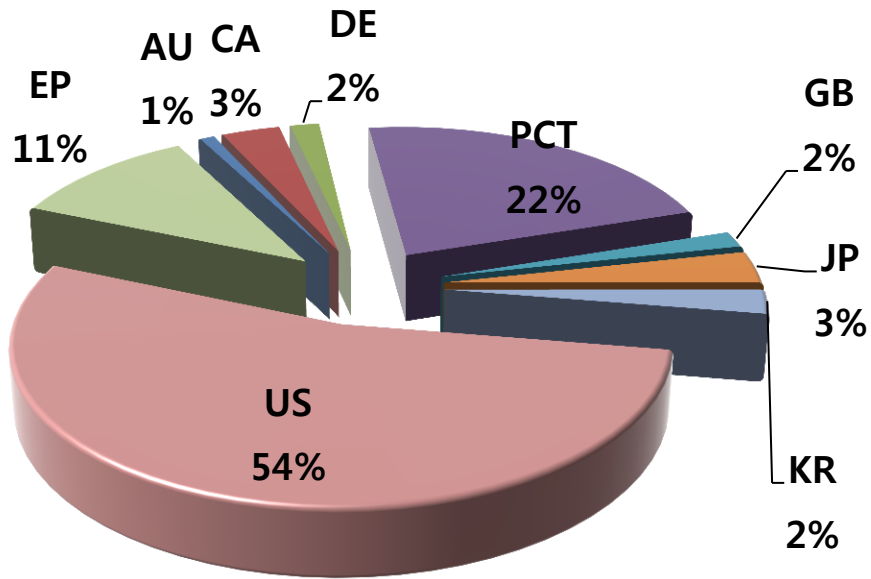
II. 보건의료부문 빅데이터 특허출원동향 (1/3)



분석조건

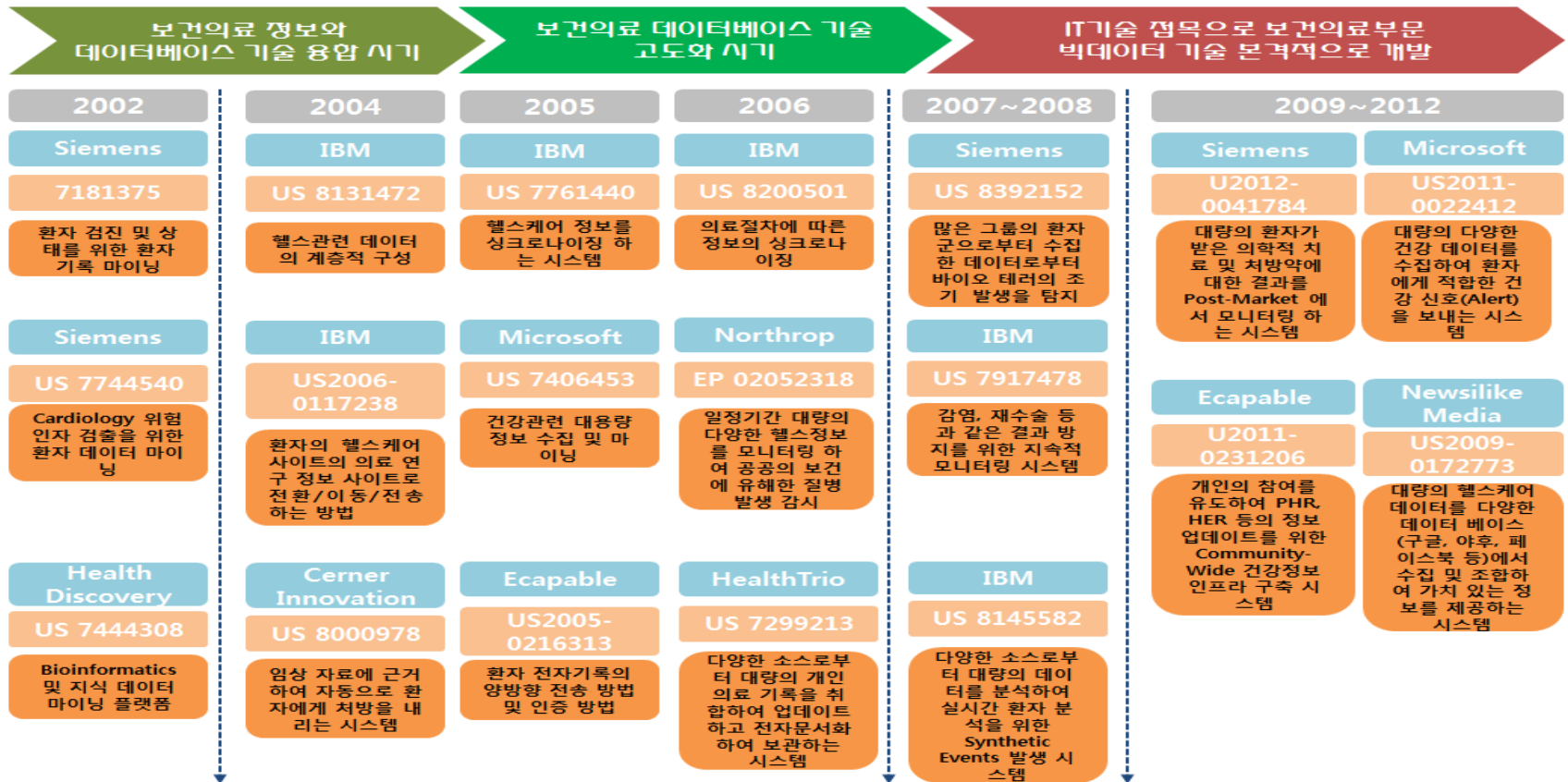
- 데이터의 규모, 데이터의 다양성 및 데이터의 처리속도 중 2가지 이상 포함한 공개특어 및 등록특어
- 1998년부터 2013년 6월까지 특허 출원일을 기준 (2012년은 이후는 미공개특어를 포함)
- 대상국가: 미국, 유럽 각국, 한국, 일본, 오스트레일리아, 캐나다, 국제출원(PCT)
- 특허DB: 포커스트 DB

보건의료부문 빅데이터 국가별 및 기업별 출원동향 (2/3)



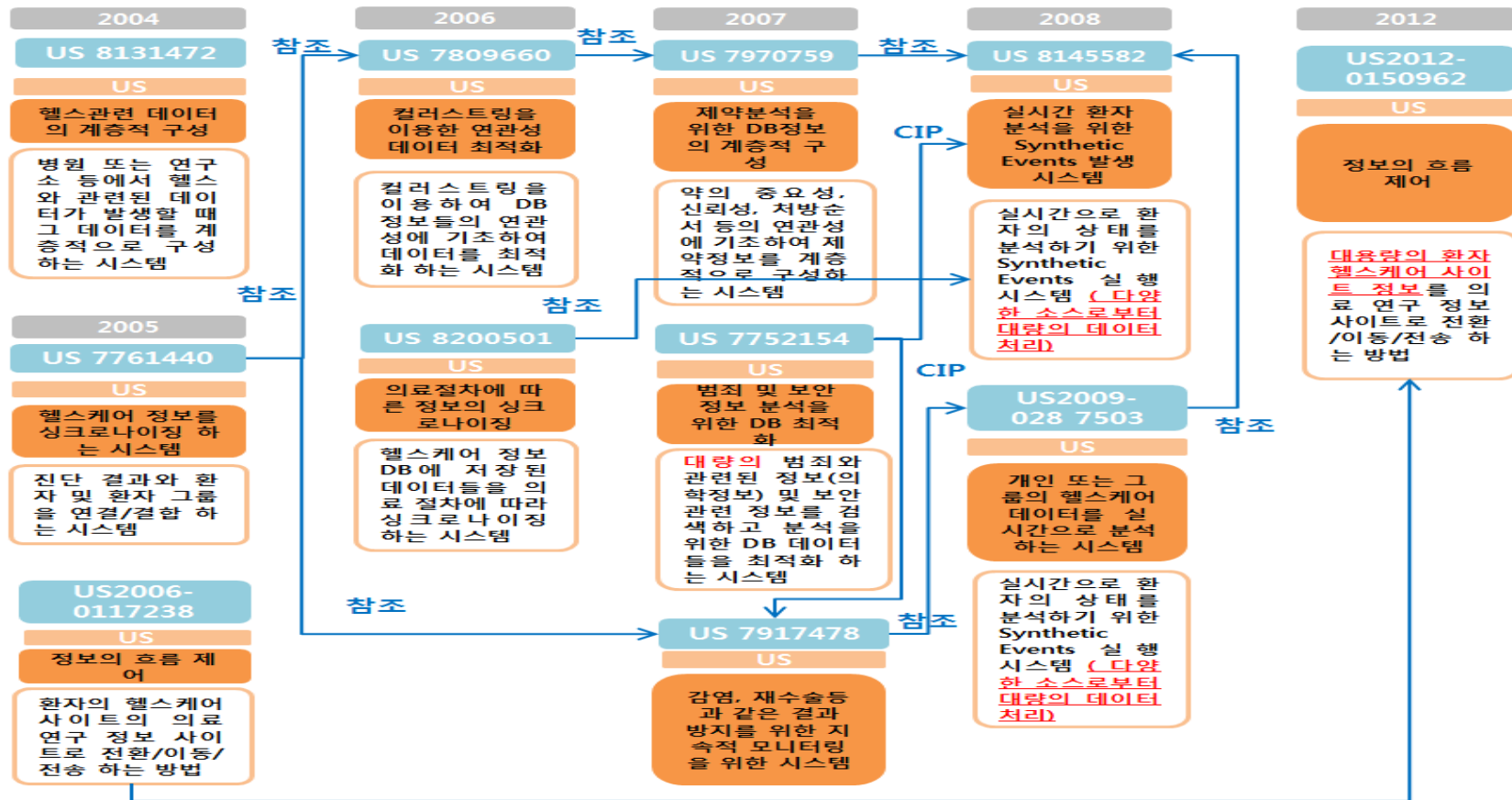
- 보건의료부문 빅데이터 특허출원 국가는 미국 특허청 출원 특허가 절반이상 차지
- 한국, 일본, EP의 출원인은 대부분 미국 국적의 출원인
- GE의 경우 1개의 특허로 한국, 일본, EP에 출원

보건의료부문 빅데이터 특허출원 분석 (3/3)



- ▶ 보건의료부문 빅데이터는 보건의료 정보를 디지털화 (데이터베이스와)하는 기술에서 출발
- ▶ 보건의료 지식(정보) + 데이터베이스 기술 + IT 기술 = 보건의료부문 빅데이터 기술(융복합 산업기술)
- ▶ 일반적으로 2007년부터 보건의료부문의 빅데이터 기술이 특허로 출원된 것을 확인할 수 있음

III. IBM 보건의료부문 빅데이터 특허출원 분석 (1/4)



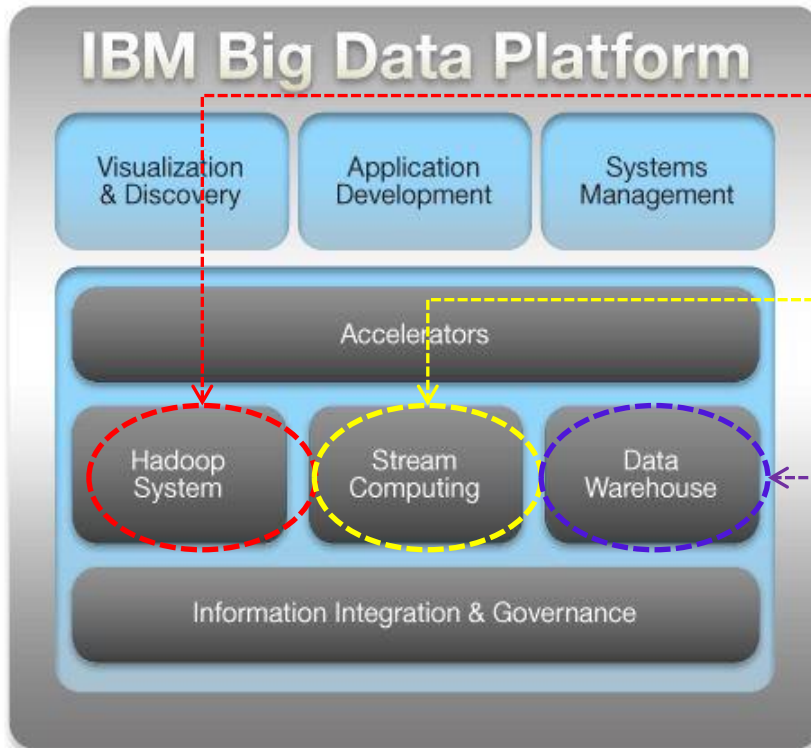
- 7명의 발명자가 4~5명씩 그룹을 지어 특허출원 (이전 특허 참조 및 CIP 출원사용)
- 자사의 데이터베이스 기술을 보건의료부문 정보처리 기술에 적용하여 보건의료부문 빅데이터 기술개발
- 데이터 마이닝, 데이터 평가(가치있는 정보 제공) 및 실시간 처리 등, 빅데이터 처리를 위한 기술 개발

IBM 보건의료부문 빅데이터 특허출원 분석 (2/4) - 저장관리부문

데이터 마이닝 기술							
비정형 정보 처리 기술							
2002	2006	2007	2008	2009	2011		2012
US 8214391	US 7836004	US 8108413	US 8280903	US 8356047	US2012-0209852	US2012-0265780	US2012-0259890
US	US	US	US	US	US	US	US
지식기반 데이터 마이닝 시스템	데이터 규칙 발견을 위한 데이터 마이닝 알고리즘 사용 시스템	비정형 자유형태의 데이터에서 자동으로 특징을 추출하는 시스템	비정형 데이터를 수집/분석/관리하는 시스템	인텔리전트 이벤트에 기반한 Unstructured 정보 데이터 마이닝	실시간 데이터 마이닝	비정형, 분산형 데이터로부터 On-Demand 방식의 연관 데이터 수집 방법	지식기반 데이터 마이닝 시스템
다수의 검색 수행자가 다양한 소스를 동시에 마이닝 가능하도록 하는 시스템. 마이닝 결과를 분류기반으로 정렬하여 기록하는 방법 사용	연관성 분석 및 트리 분류 방식을 사용하여 마이닝한 데이터의 규칙을 발견하는 시스템으로 사용자 인터페이스를 통하여 사용자가 최초 규칙을 설정할 수 있도록 사용자 인터페이스를 제공하는 사용자 설정 데이터 규칙 발견 시스템	비정형 데이터 및 지정된 위치에 고정된 데이터에 기계학습 및 연관성 분석을 사용하여 언어 모델 레이블을 부여하여 특징을 추출하는 시스템	UIMA(Unstructured Information Management Architecture)에 관한 것으로 비정형 데이터의 처리도 함께 가능하게 하는 시스템	Unstructure 정보를 검색하는데 있어서, 검색어에 따라 적절한 검색엔진을 제공하고 검색어 및 검색결과에 가중치를 주어 가치있는 검색 결과를 제공하는 시스템	실시간으로 사용자의 오피니언, 경향 또는 인터넷 사용 습관 등을 실시간으로 마이닝하는 시스템	분산형의 비정형 데이터 소스로부터 연관성 분석 엔진을 사용하여 on-demand 방식으로 연관성이 있는 데이터를 수집하는 방법을 제공하는 시스템	다수의 검색 수행자가 다양한 소스를 동시에 마이닝 가능하도록 하는 시스템. 마이닝 결과를 검색어 분류기반으로 정렬하여 기록하는 방법 사용 - Unstructured 데이터도 처리 가능한 기능 추가
2004	2010		2010	2010		2013	2012
US 8307109	US 8326669	US 8103621	US 8326669	US 8103621	US2012-0246154	US2013-0086116	US 8538988
US	US	US	US	US	US	US	US
기업규모 데이터의 실시간 통합 서비스 시스템	자동저장관리를 위한 교정(Corrective Action) 선택 및 스케줄링	대량의 파일을 계층적으로 저장 관리 하는 시스템에서 파일을 복원하는 방법	대량의 파일(Billions of files)을 계층적으로 관리하는 파일 관리 시스템에서 훼손된 파일 또는 파일의 일관성을 유양을 지하기 위한 방향 복원방법을 제공하는 방법	소셜 네트워크 연결자 간에 다수의 클라우드 소스 공유	연관성에 기반한 검색결과 수집	데이터의 병렬처리를 위한 작업흐름에 따른 데이터 통합	상호작용 데이터 마이닝을 위한 선택적 마이닝 모델 저장
기업간 데이터를 실시간으로 통합하는 서비스 시스템으로 Batch-mode 도 함께 서비스 할 수 있도록 설계되어 있고, 이런 서비스는 사용자 인터페이스를 통하여 간편하게 접근할 수 있도록 설계 되어 있음	MAPE방식을 사용하여 데이터 저장관리를 자동으로 하는 시스템으로 데이터 저장 시스템의 에러(Violation)를 M(모니터링), A(분석), P(계획) 및 E(실행)를 사용하여 실시간으로 저장관리의 교정(Corrective Action)을 수행하는 시스템	대량의 파일(Billions of files)을 계층적으로 관리하는 파일 관리 시스템에서 훼손된 파일 또는 파일의 일관성을 유양을 지하기 위한 방향 복원방법을 제공하는 방법	각 소셜 네트워크에 연결된 클라우드 소스를, 소셜 네트워크간의 연결을 통하여 클라우드 소스도 함께 공유할 수 있게 하는 클라우드 소스 관리 시스템	다 수 의 Heterogeneous 소스로부터 검색어에 대한 결과를 연관성에 기반하여 수집하는 시스템	Structured, Semi-Structured 및 Unstructured 데이터를 작업 흐름에 따라 통합하여 데이터를 병렬 처리 방식으로 처리하는 방법에 대한 시스템	데이터 마이닝의 모델을 각 마이닝의 특징에 따라 선택적으로 사용하고, 사용된 특징적 마이닝 모델을 각기 다른 장소에 저장하는 방법에 관한 시스템	

- ▶ 빅데이터 저장관리 기술을 보건의료부문 빅데이터 기술에 적용한 것을 알 수 있음
- ▶ 저장관리 부문에서 비정형 정보 처리 기술을 특허출원한 2007년부터 보건의료부문 빅데이터 관련특어를 출원
- ▶ 결론적으로, **데이터 베이스 기술력 확보가 향후 보건의료부문 빅데이터 기술력 확보와 직결된다는 것을 알 수 있음**

IBM 빅 데이터 처리 플랫폼 구성 (3/4)



1. 아이비엠 자체 아둠 시스템

- 적응적 맵리듀스(Adaptive MapReduce) 적용
- 아둠 시스템 스줄러 적용

2. 입력 스트리밍 컴퓨팅 적용

- 입력 스트리밍 병렬처리(실시간 처리용)
- 고 확장성

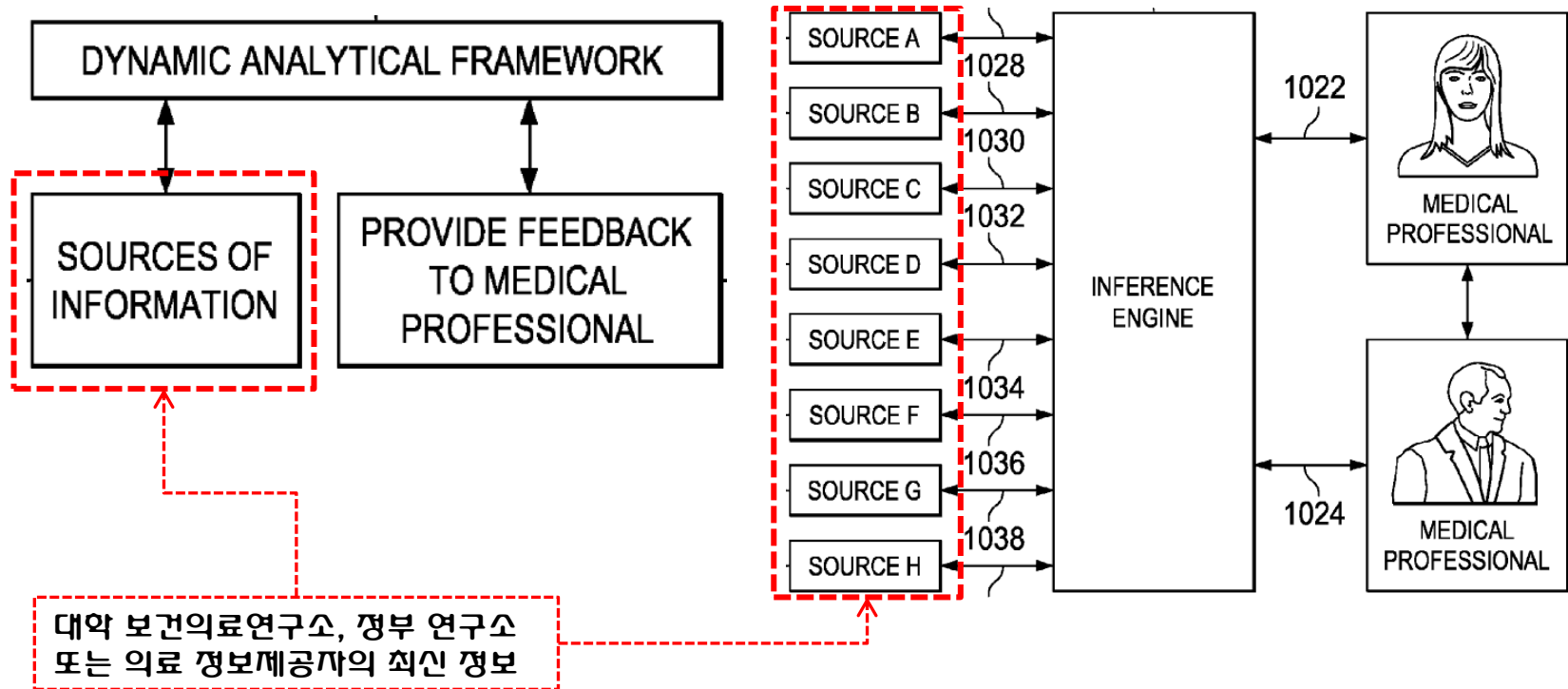
3. 데이터 웨어하우스

- 고성능 병렬프로세서 엔진 적용
- 고성능 OLAP (Online Analytic Processing) 적용

- 아이비엠 빅 데이터 처리 플랫폼은 표준에서 제공하는 오픈소스 아둠(HDFS, MapReduce 등)과 쿼리 언어로 Jaql, Pig 및 Flume 등을 포함하고 있음
- 또한, 데이터 관리를 위한 Hive, 텍스트 서치를 위한 Lucene 등을 포함하고 있음

IBM 보건의료부문 빅데이터 특허분석 (4/4)

- IBM 특허 US 8,145,582 분석
- 발명의 명칭: Synthetic Events for Real Time Patient Analysis



- 네트워크로 연결된 보건의료부문 최신정보들을 통하여 환자에게 최신의 의료정보를 **실시간**으로 제공하는 시스템
- 환자로부터 받은 정보와 최신의 의료정보를 **연관성**에 근거하여 비교하고 그 중 가능성이 가장 높은 정보 순으로 제공

IV. 보건의료부문 빅데이터 적용 사례 - 미국



- 미국은 의료부문에 있어 ①제약회사 R&D, ②영상 데이터, ③의료배상 청구 및 가격 및 ④환자행동 및 감정 데이터의 4개 부문에 의료 빅 데이터(Major Pool of Data) 적용
- 현재는 각 섹터 안에서만 데이터들이 공유되지만, 각 섹터간에 데이터 풀의 통합 요구가 커지고 있음 (의료기록 유출에 의한 법률적 문제 및 의료 데이터 전산화 문제 있음)

빅 데이터 적용 성공사례 – 미국/유럽연합

	기관	성공 사례 내용
미국	The Department of Veterans Affairs(VA)	헬스 케어 정보(HIT, Health Information Technology) 및 원격 환자 모니터링 시스템(Remote Patient Monitoring System)을 성공적으로 시행함 임상 가이드라인에 따른 증거 기반(Evidence-base)의 약물 치료에 높은 성공률을 보임
	Kaiser Permanente Consortium (캘리포니아 기반 통합 케어 카이저 퍼머난테 컨소시움)	임상 데이터와 약의 가격 정보를 연결하여, 바이옥스(Viox: 치료시 위험 부담이 제거된 상태로 통증과 염증을 완화시키는 Cox-2 인자)의 약리 부작용을 발견하고 시장으로부터 약을 퇴출시키는데 역할을 하는 데이터 베이스를 제공하는데 성공함
유럽연합	The National Institute for Health and Clinical Excellence (영국)	새로운 약의 가격대비 효과에 대한 정보 및 현재 사용되는 고가의 치료방법에 대한 대용량 정보의 임상 데이터베이스를 사용하여, 국가 헬스 서비스에 새로운 약에 대한 가이드라인을 제시함
	The Italian Medicines Agency (이탈리아)	국가 가격대비 효과 프로그램(National cost-effectiveness program)의 한 부분으로 고가의 새로운 약의 임상 데이터를 수집하고 분석하여, 새 약에 대한 조건적 비용변제 등급을 도입하여, 임상 데이터 연구의 결과를 반영한 시장-접근적 조건 및 가격을 재평가하는 일을 함

감사합니다!

異想

특허법인 이상
E-Sang Patent & Trademark Law Firm

서울시 서초구 양재동 82-2 우도빌딩 3층, 5층
02) 2057-1246~7
esang@espat.co.kr