

# Big Data Curation & Its Application



Hanmin Jung  
Head of Dept. of Computer Intelligence Research  
KISTI

# Let Me Introduce Myself :-)

## 인물 정보



**정한민** 연구원, 교수

소속 [한국과학기술정보연구원](#) (실장), [과학기술연합대학원대학교](#) (겸임교수), [한국외국어대학교](#) (초청연구원)

학력 [포항공과대학교 대학원 컴퓨터공학 박사](#)

[포항공과대학교 대학원 전자계산학 석사](#)

경력 [한국과학기술정보연구원 소프트웨어연구실 실장](#)

[과학기술연합대학원대학교 응용정보과학 겸임교수](#)

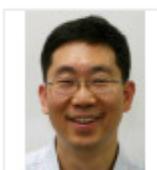
[한국콘텐츠학회 편집위원, 미사](#)

[한국정보과학회 컴퓨터지능소사이어티 이사](#)

[프로필 더보기](#)

도서 [IT에 의한 뉴 비즈니스 ...](#) 미담북스 2011.06.30

## 뉴스 정확도 | 최신순



[\[과학세상/정한민\] 빅 데이터 시대의 '데이터 큐레이션'](#)

동마일보 | A31면3단 | 2012.05.02 (수) 오전 3:20

정보에 대한 최종 필터링은 결국 사용자의 둘이라는 사실은 변함이 없으므로, 스스로의 분별 능력을 키우려는 노력을 게을리 해서는 안 될 것이다. **정한민** 한국과학기술정보연구원 소프트웨어연구실 실장 네이버에서 보기 | 관련기사 보기 | 이 언론사 내 검색



[정한민 KISTI연구원 세계인명사전 그랜드 슬램 달성](#) 전자신문 | 2011.09.14 (수) 오후 3:48

정한민 한국과학기술정보연구원(KIST) 책임연구원이 세계 3대 인명사전에 모두 등재되는 '인명사전 그랜드슬램'을 달성했다. 정 연구원은 최근 미국 인명정보기관(ABI)으로부터 '세계의 뛰어난...' 네이버에서 보기 | 관련기사 보기 | 이 언론사 내 검색

# Let Me Introduce Myself :-)

---

## ■ Recent Social Activities (2012~)

- 보건복지부 빅데이터 기술 전문가 패널
- 한국연구재단 융합연구우수성평가위원회 위원
- 국립전파연구원 방송통신표준 전문위원
- 기술표준원 빅데이터 표준화 프레임워크 표준연구회 위원
- ITRC 과제기획위원회 빅데이터 분과, UI/UX 분과 기획위원/기술간사
- 교육과학기술부 과학기술 빅데이터 포럼 전문가위원회 간사
- 국가경쟁력강화위원회 ‘범부처 Giga KOREA 사업’ 전문가 위원회 위원
- 국가과학기술위원회 기술영향평가위원회 위원
- 방송통신위원회 빅데이터 포럼 인력양성분과 위원장/자문위원/운영위원
- IBS 장비심의위원회, 지식경제부 중앙장비심의위원회 심의위원
- 한국과학창의재단 과학창의앰배서더
- 지식경제부 ‘SW+인문 포럼’ 멘토
- IT VC 전문성 강화 프로그램 전문강사
- ISO/IEC JTC1/SC32 전문위원회 위원, JTC 1/SC34 표준개발위원회 위원
- ISO 20022 TC68 Advisory Expert
- 한국정보통신기술협회 정보통신표준화위원회 –  
디지털콘텐츠/SW품질평가/메타데이터/사물지능통신 프로젝트 그룹 위원
- 교육과학기술부 차세대정보컴퓨팅분과 – 기획위원
- 국무총리실 정보지식인대회 출제/평가위원
- 문화체육관광부 공공문화정보 전문가 위원

# Trend Reports

---

## ■ Weekly IT Trends (NIPA)

- 과학기술 빅데이터 동향 및 활용방안 (Vol. 1593)
- 모바일 비즈니스 인텔리전스 기술 동향 (Vol. 1573)
- 시맨틱 소셜 네트워크를 구성하는 온톨로지 어휘 기술 현황 (Vol. 1560)
- 정부 빅데이터 분석 기술 적용 동향 (Vol. 1556)
- 제품 및 기술 수명 주기 활용 동향 (Vol. 1550)
- 개체 식별 관점에서 바라본 링크드 데이터 동향 (Vol. 1524)
- 유망 기술 발굴 모델 및 서비스 동향 (Vol. 1521)
- 포지셔닝의 측면에서 살펴본 모바일 태블릿 컴퓨터 동향 (Vol. 1486)
- 테크놀로지 인텔리전스 동향 (Vol. 1477)
- 추론 기술 연구 동향 (Vol. 1446)
- 트리플 레파지토리 벤치마킹 (Vol. 1439)
- 차세대 IT 기기와 HCI 기술 동향 전망 (Vol. 1435)
- 시맨틱 검색 기술 동향 (Vol. 1431)
- 시맨틱 웹 국내 특허 동향 (Vol. 1420)
- 감성 분석과 브랜드 모니터링 기술 동향 (Vol. 1396)
- 시맨틱 웹이 경제·사회에 미치는 영향 (Vol. 1372)
- 웹 매핑 서비스 비교 분석 (Vol. 1352)
- 시맨틱 웹 2.0 기술 동향 (Vol. 1344)
- 국내 포털 검색 시장 및 특허 동향 (Vol. 1341)
- Open API 기술 동향 (Vol. 1296)
- 엔터프라이즈 검색 기술 동향 (Vol. 1276)
- 전자상거래 검색 기술 동향 (Vol. 1273)
- 시맨틱 웹 포털 기술 동향 (Vol. 1264)

# Art Curator

---



<http://www.chrysler.org/files/resources/gallery-talk-3.jpg>

# Press

東亞日報

| 과학세상 |



정한민

한국과학기술정보연구원  
소프트웨어연구실 실장

큐레이터는 박물관과 미술관에서 재정 확보, 유물의 보존 관리, 자료 전시, 홍보활동을 하는 사람을 일컫는다. 그럼 큐레이터란 직업은 예술 분야에만 있는 것일까? 그렇지 않다. 2011년 2월 세계적 학술지인 네이처는 생물정보학 분야에서 생물정보 데이터베이스의 내용을 검토하고 주석을 달아 완성도를 높이는 '생물정보 큐레이터'라는 직업을 소개했다.

'데이터 큐레이션'이라는 키워드가 뜨고 있다. 위의 정의로 짐작되듯 데이터를 발굴하고, 검색하며, 품질을 유지하면서 가치를 부여하는 활동이다. 데이터 큐레이션이 구체적으로 무엇을 의미하며, 왜 화두가 되고 있는 것일까? '빅 데이터'라는 요즘 가장 '핫한' 키워드를 가지고 풀어보기로 하자.

한국정보화진흥원에 의하면 2011년 전 세계 디지털 정보량은 약 1.8ZB(제타바이트·1ZB는 1조 GB·기가바이트)이며, 2020년 관리해야 할 정보의 양은 현재의 50배 이상 증가할 것으로 예상된다. 이렇듯 거대한 양으로, 빠른 속도로 증가하는 빅 데이터 환

경에서 우리는 정보를 찾는 데 점점 지쳐가고 있다. 사용자가 직접 키워드를 입력해서 원하는 정보를 찾는 전통적인 정보 검색 방식에서는 사용자가 직접 큐레이터의 역할까지 겸했다. 그러나 엔지니어들조차도 정보를 검색하고 수집하는 데 근무시간의 절반 가까이 할애한다는 연구 결과에서 보듯 개인에게 큐레이션의 역할까지 맡기기에는 한계가 너무나 명확한 게 현실이다.

'데이터 큐레이션'은 빅 데이터의 속성을 고스란히 갖고 있는 소셜네트워크서비스나 소셜미디어, 스마트폰에 내장된 각종 센서, 웹 데이터에서 정제된 정보를 찾고자 하는 기법이다. 용어는 생소하지만 예전부터 다양한 영역에서 데이터 큐레이션 기법이 쓰여 왔다. 예를 들어 새우깡 제품 속에서 생쥐 머리가 발견됐을 때의 충격을 계기로 국내 주요 기업들은 '고객의 소리'를 수집하고 정제된 정보를 걸러내 루머나 사태가 확산되기 전에 조치할 수 있도록 하는 방안을 몇 년 전부터 도입하여 운영하고 있다. 인터넷 미디어에서도 데이터 큐레이션 기법이 쓰인다. 미국 버락 오바마 대통령 당선에 영향을 끼친 것으로 잘 알려진 블로그 뉴스미디어인 '허핑턴포스트'는 필진의 지명도와 신뢰를 바탕으로 한 달 순방문자 수가 수천만 명에 이를 정도로 그 영향력을 과시하고 있다.

신문 레이아웃을 이용하여 관심 정보를 자동 편집해서 제공하는 페이퍼닷리(Paper.li), 개인화된 인터넷 잡지로서의 라이브스탠드(Livestand) 등도 데이터 큐레이션의 효과를 확실하게 보여준다.

최근 트위터와 같은 소셜미디어의 엄청나게 빠른 전파 속도로 인해 확인되지 않은 각종 루머의 난무로 이른바 '쓰레기 정보'를 정화하려는 노력이 커지고 있다. 이에 따라 자신만의 방식으로 데이터를 해석하고 중요도순으로 배치하여 친구나 추종자에게 제공하는 콘텐츠 유통자로서의 '소셜 큐레이터'들이 등장했는데 이는 웹 2.0 시대에 '프로슈머'가 소셜네트워크상에서 진화한 모습이라고 할 수 있다.

그렇지만 우리가 주의하고 유념해야 할 것은 사용자가 추구하는 가치와 데이터 큐레이션에 담겨진 의도 사이에서의 균형 감각이다. 가치에 치우치다 보면 특정 분야에 대한 집중도와 방향성이 떨어질 수 있고, 의도에 치우치면 '그들만의 리그'로 전락할 위험이 있다. 특히 불순한 의도가 담긴 큐레이션은 결국 또 다른 쓰레기 정보를 생산할 수 있다는 점을 잊지 말아야 한다. 정보에 대한 최종 필터링은 결국 사용자의 몫이라는 사실은 변함이 없으므로, 스스로의 분별 능력을 키우려는 노력을 게을리 해서는 안 될 것이다.

## 빅 데이터 시대의 '데이터 큐레이션'

# Data Curation

---

## ■ Definition

- Active management of data over its lifecycle
- Ensuring that data is trustworthy, discoverable, accessible, reusable, and fit for use
  - E.g. data preparation for analytics

## ■ Curation Types

- Manual curation
- (Semi-)automatic curation
  - E.g. data cleansing, record duplication, classification
- Sheer curation (curation at source)
  - Integrated in workflow for creating and managing data

E. Curry, A. Freitas, and S. O'Riain, "Data Curation at the New York Times", 2011.

# Data Curation

---

## ■ Steps to Setup the Process

- Identify what data you need to curate
- Identity who will curate the data
- Define the curation workflow
- Identify appropriate data-in & data-out formats
- Identity the artifacts, tools, and processes needed to support the process

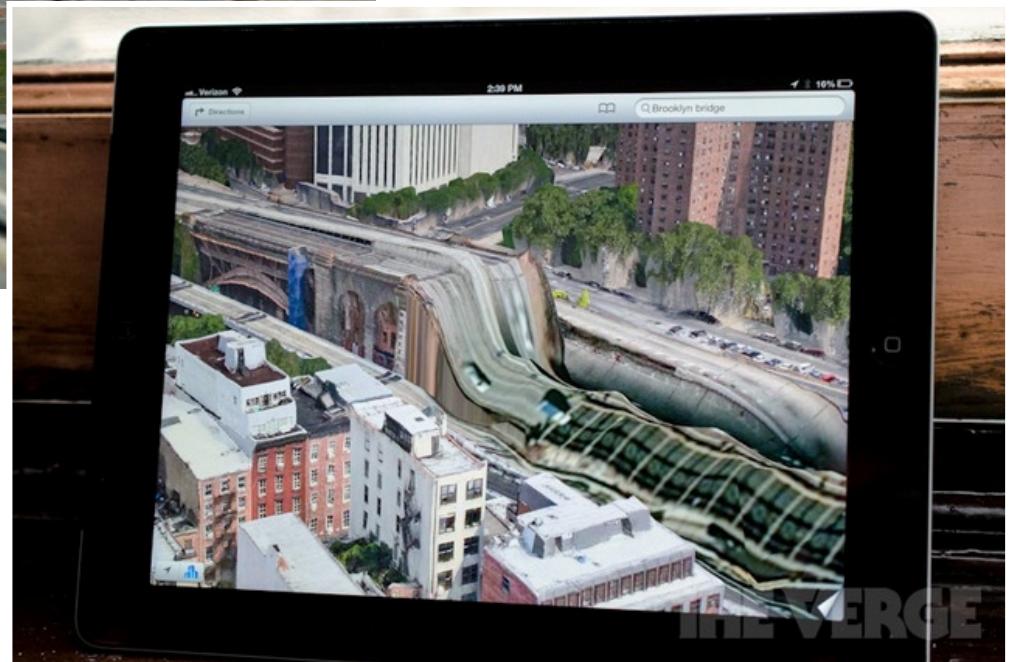
# Case Studies

---

- 1. Apple Maps*
- 2. Strategic Foresight*
- 3. Big Data*

# Apple Maps

---



[http://cdn0.sbnation.com/entry\\_photo\\_images/5667445/20120920-DSC\\_7192VERGE\\_large\\_verge\\_medium\\_landscape.jpg](http://cdn0.sbnation.com/entry_photo_images/5667445/20120920-DSC_7192VERGE_large_verge_medium_landscape.jpg)

# Google Maps vs. Apple Maps

Category	Winner	Why
Street Layout	Google Maps	More traditional, easier to use and to see
Traffic Details	Google Maps	Better and more accurate traffic details, available in more places
Speed	Tie	Roughly the same
Power Consumption	Tie	You couldn't tell the difference
Ease of Use (iOS)	Apple Maps	Works across all applications natively
Non-driving Features	Google Maps	Walking, biking, and public transportation directions built in
Voiced Turn-by-Turn	Google Maps	Overall better performance, from voice quality to timing
Additional Features	Apple Maps	Full access to your phone means instant contact driving details, better notification features, and overall more convenient features
Design	Google Maps	So simple yet so finely tuned
Overall Winner	Google Maps	7 - 4

# Apple Siri



## Choose a movie.

Ask Siri to get showtimes, look up movie facts, play trailers, show you reviews, and more.

## Find a restaurant.

Ask Siri to search by different criteria or a combination. Siri gets you photos, reviews, and reservations.

<http://www.apple.com/ios/siri/>

# We can see it as much about it as we know

## ■ Google's Financial Table

Revenues	Full Year			2012			
	2010	2011	(unaudited)	Q1	Q2	(unaudited)	Q4
			2012			Q3	
Google Websites	\$19,444	\$26,145	\$31,211	\$7,312	\$7,542	\$7,727	\$8,640
Y/Y Growth Rate	24%	34%	19%	24%	21%	15%	18%
Q/Q Growth Rate	NA	NA	NA	0%	3%	2%	12%
Google Network Members' Websites	\$8,792	\$10,386	\$12,465	\$2,913	\$2,983	\$3,133	\$3,436
Y/Y Growth Rate	23%	18%	20%	20%	20%	21%	19%
Q/Q Growth Rate	NA	NA	NA	1%	2%	5%	10%
Total Advertising Revenues	\$28,236	\$36,531	\$43,686	\$10,225	\$10,525	\$10,860	\$12,076
Y/Y Growth Rate	23%	29%	20%	23%	21%	16%	19%
Q/Q Growth Rate	NA	NA	NA	1%	3%	3%	11%
Other Revenues	\$1,085	\$1,374	\$2,354	\$420	\$439	\$666	\$829
Y/Y Growth Rate	42%	27%	71%	56%	42%	73%	102%
Q/Q Growth Rate	NA	NA	NA	2%	5%	52%	24%
Total Google Revenues	\$29,321	\$37,905	\$46,039	\$10,645	\$10,964	\$11,526	\$12,905
Y/Y Growth Rate	24%	29%	21%	24%	21%	19%	22%
Q/Q Growth Rate	NA	NA	NA	1%	3%	5%	12%

<http://investor.google.com/financial/tables.html>

# Nokia's Burning Ships Strategy



[http://www.brightsideofnews.com/Data/2011\\_5\\_5/Steven-Elop-Burn-Boats-Strategy-is-a-Cortez-Move-for-Nokia/Hernando\\_Cortez\\_BurningBoat.jpg](http://www.brightsideofnews.com/Data/2011_5_5/Steven-Elop-Burn-Boats-Strategy-is-a-Cortez-Move-for-Nokia/Hernando_Cortez_BurningBoat.jpg)

# Result of the Strategy

## ■ Worldwide Market Share

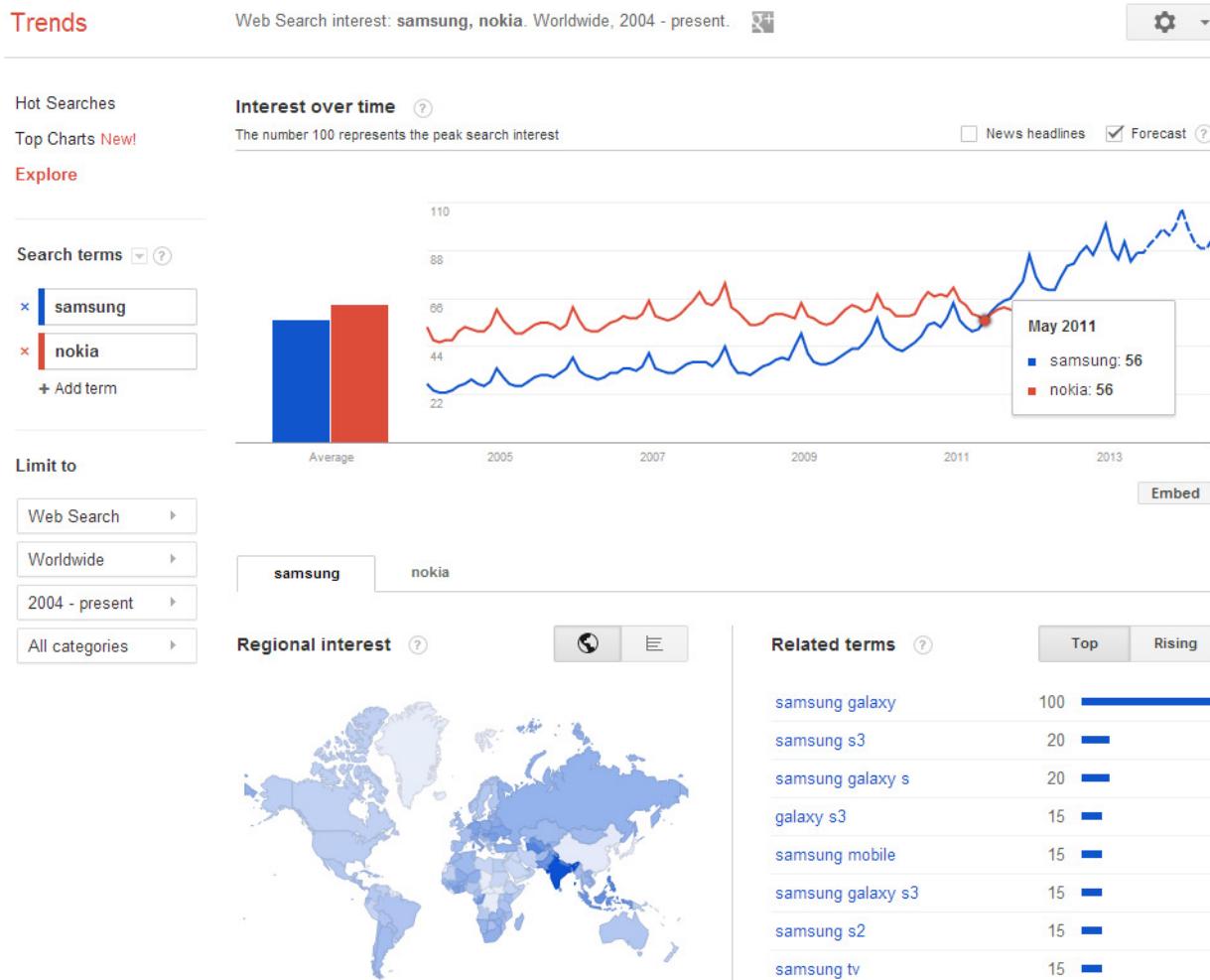
- Worldwide mobile device sales to end users in 2008 ~ 2013

Company	1Q2013 (%, M. Units)	3Q2012 (%, M. Units)	3Q2011 (%, M. Units)	3Q2010 (%, M. Units)	3Q2009 (%, M. Units)	3Q2008 (%, M. Units)
Samsung	27.5, 115.0	23.7, 105.4	22.3, 87.8	20.5, 71.4	21.0, 60.2	17.0, 52.0
Nokia	14.8, 61.9	18.7, 82.9	27.1, 106.6	31.6, 110.4	37.8, 108.5	38.6, 117.9
Apple	8.9, 37.4	6.1, 26.9	4.3, 17.1	4.0, 14.1		
LG	3.7, 15.4	3.1, 14.0	5.4, 21.1	8.1, 28.4	11.0, 31.6	7.5, 23.0
ZTE	3.2, 13.5	3.1, 13.7	4.9, 19.1	3.5, 12.1		
Sony Ericsson					4.9, 14.1	8.4, 25.7
Motorola					4.7, 13.6	8.3, 25.4
Others	41.9, 175.4	45.3, 201.6	36.1, 142	32.2, 112.5	20.6, 59.1	20.1, 61.5
<b>Total</b>	<b>418.6</b>	<b>444.5</b>	<b>393.7</b>	<b>348.9</b>	<b>287.1</b>	<b>305.4</b>

Gartner, IDC Worldwide Mobile Phone Tracker

# Sign of the Result

## ■ Google Insights for Search



<http://www.google.com/insights/search/>

# Strategic Foresight

## Technology Intelligence

Identification, assessment and usage of information on emerging technologies and technological discontinuities

## Political Environment Foresight

Identification, assessment and usage of information on legislation, the political environment and on shifts in political will

## Strategic Foresight

Assessment of competitors and identification and assessment of products and service in development or available in lead markets

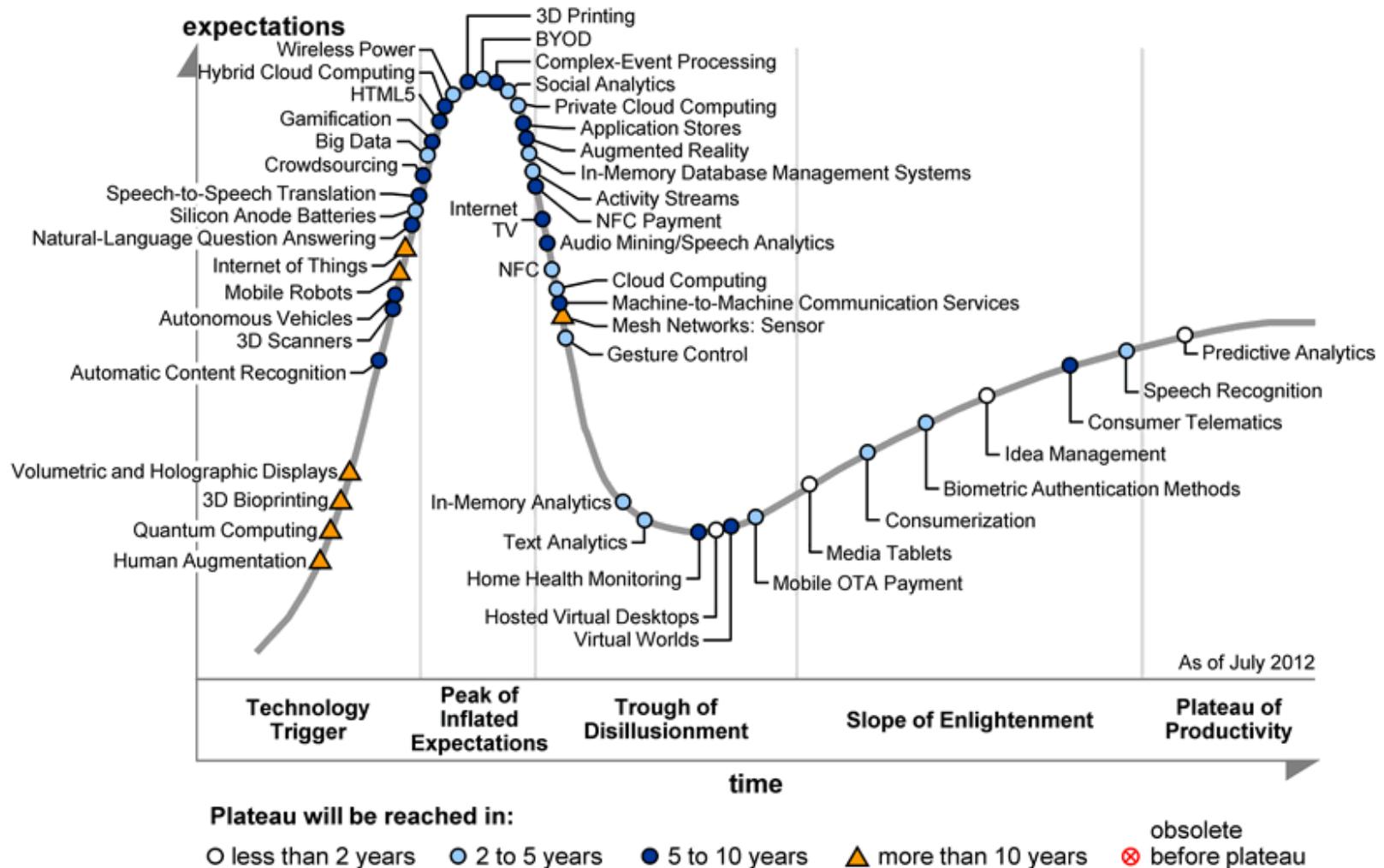
## Competitive Intelligence

Identification, assessment and anticipation of consumer needs, lifestyle and socio-cultural trends

## Consumer Foresight

# Hype Cycle

## Emerging Technologies Hype Cycle 2012



# Social Data



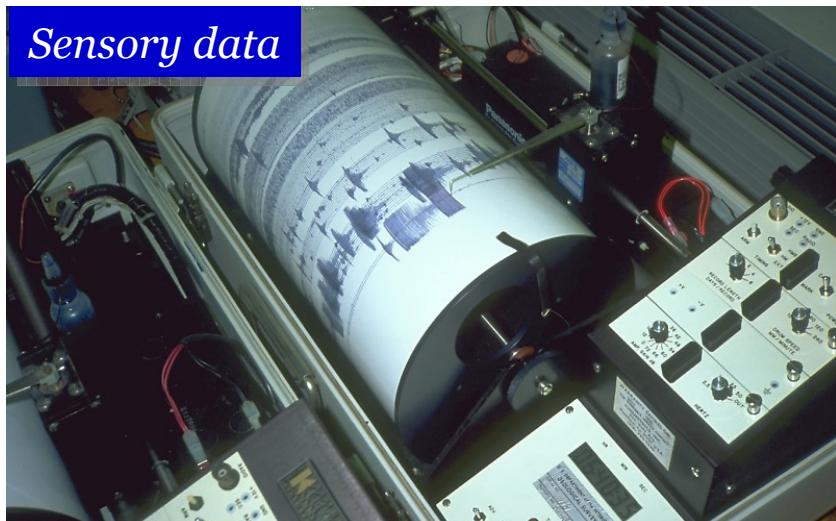
<http://bynay.files.wordpress.com/2011/08/united-noy-weblife-60-seconds.jpg>

# Machine Data

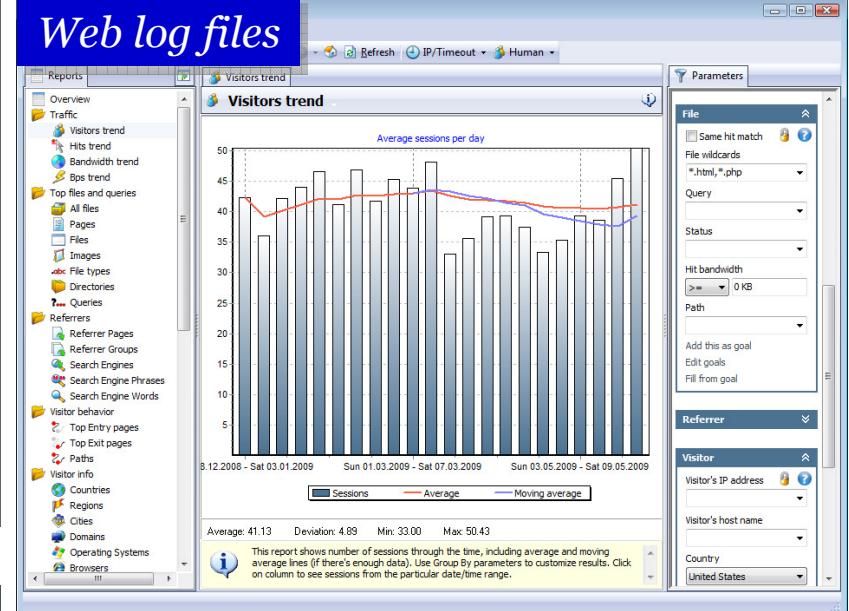
*Call data records*

	Called	Start	Released	Duration	Rel Code
ATTCARD01	111116111506-3	New York(2:21)	3016242223	11/16/2011 11:16:58	00:01:34 Normal
ATTCARD01	111116111506-24	New York(2:22)	3017242222	11/16/2011 11:15:24	00:11:05 00:00:41 Normal
ATTCARD01	111116111506-2	New York(2:11)	3016042222	11/16/2011 11:15:29	00:00:42 Normal
ATTCARD01	111116111506-21	New York(2:20)	3012242220	11/16/2011 11:15:25	00:00:41 Normal
ATTCARD01	111116111506-19	New York(2:19)	3017242218	11/16/2011 11:15:25	00:00:43 Normal
ATTCARD01	111116111506-18	New York(2:17)	3015242217	11/16/2011 11:15:25	00:00:43 Normal
ATTCARD01	111116111506-17	New York(2:16)	3016242216	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-16	New York(2:15)	3016242215	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-15	New York(2:14)	3016242214	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-14	New York(2:13)	3016242213	11/16/2011 11:15:24	00:00:43 Normal
ATTCARD01	111116111506-13	New York(2:12)	3016242212	11/16/2011 11:15:24	00:00:43 Normal
ATTCARD01	111116111506-12	New York(2:11)	3016242211	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-11	New York(2:10)	3016241010	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-10	New York(2:9)	3019242286	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-9	New York(2:8)	3019242288	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-8	New York(2:7)	3016242237	11/16/2011 11:15:24	00:00:43 Normal
ATTCARD01	111116111506-7	New York(2:6)	3019242236	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-6	New York(2:5)	3016242235	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-5	New York(2:4)	3019242234	11/16/2011 11:15:24	00:00:42 Normal
ATTCARD01	111116111506-4	New York(2:3)	3016242233	11/16/2011 11:15:24	00:00:41 Normal
ATTCARD01	111116111506-22	New York(2:21)	3019242221	11/16/2011 11:15:25	00:00:41 Normal
ATTCARD01	111116111506-20	New York(2:19)	3012242219	11/16/2011 11:15:25	00:00:41 Normal

*Sensory data*



*Web log files*

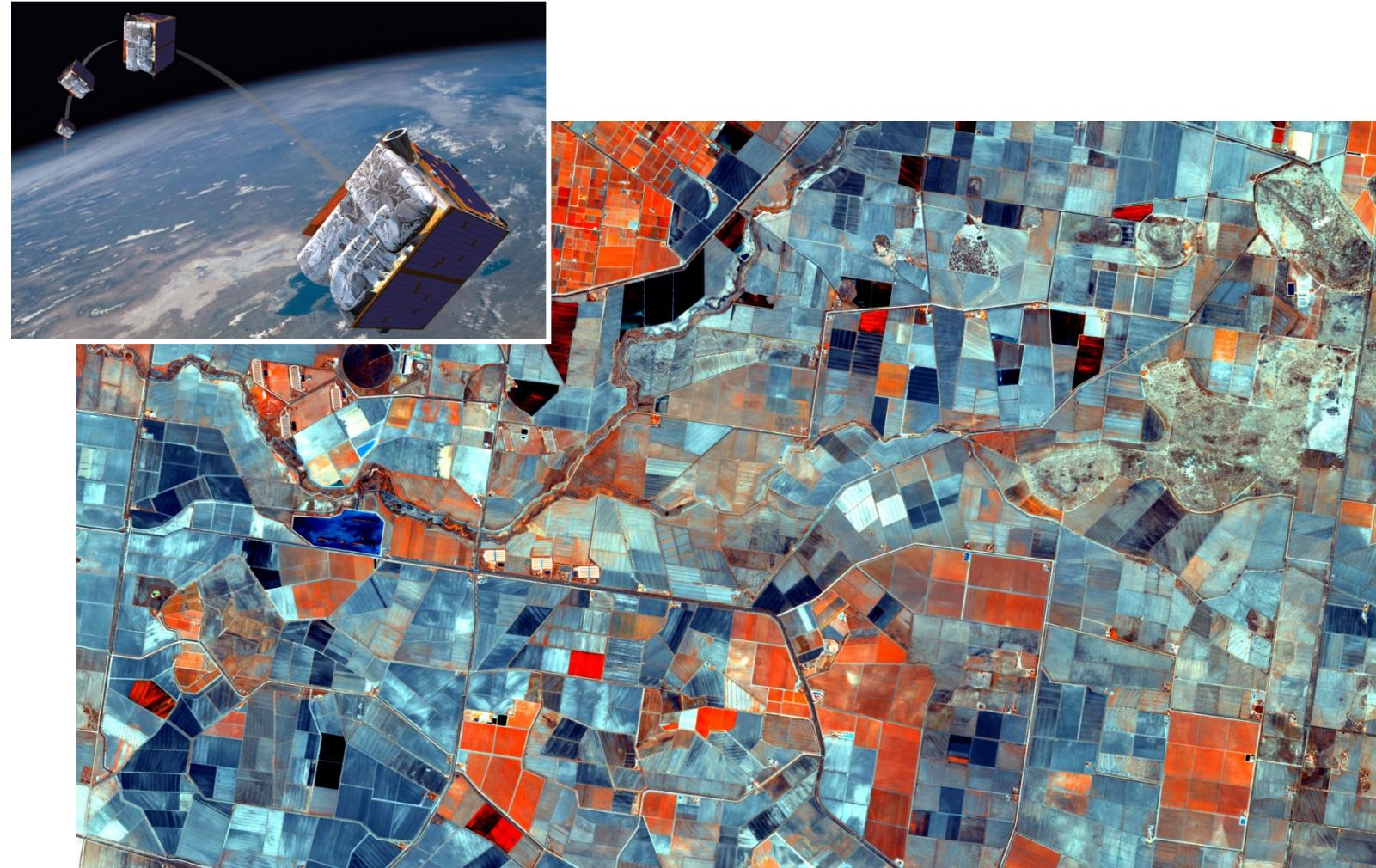


*Financial Instrument Trade*



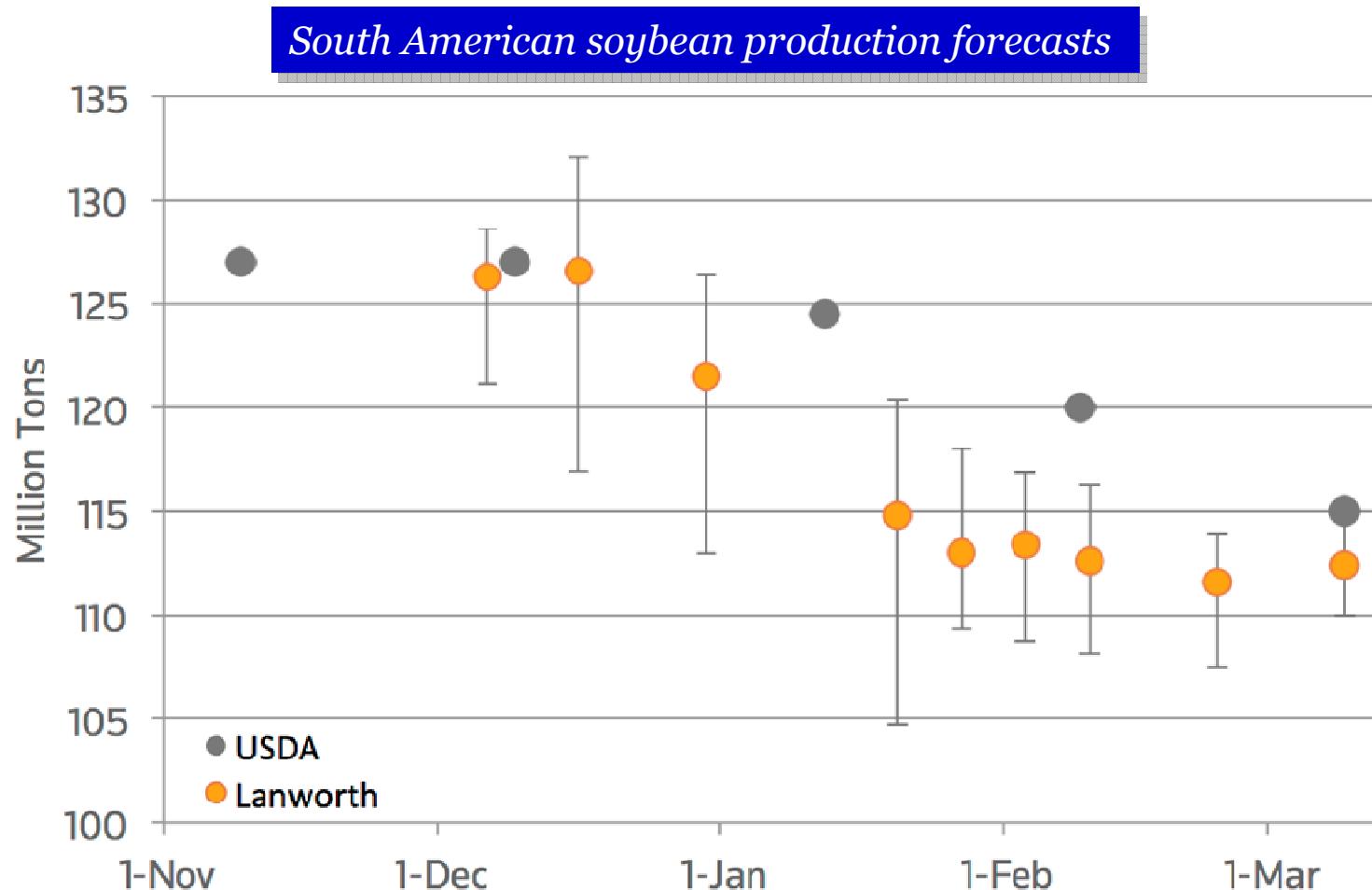
T. Baer, "What is Big Data? The Reality for Analytics", OVUM, 2011.

# Crop Yield Estimation

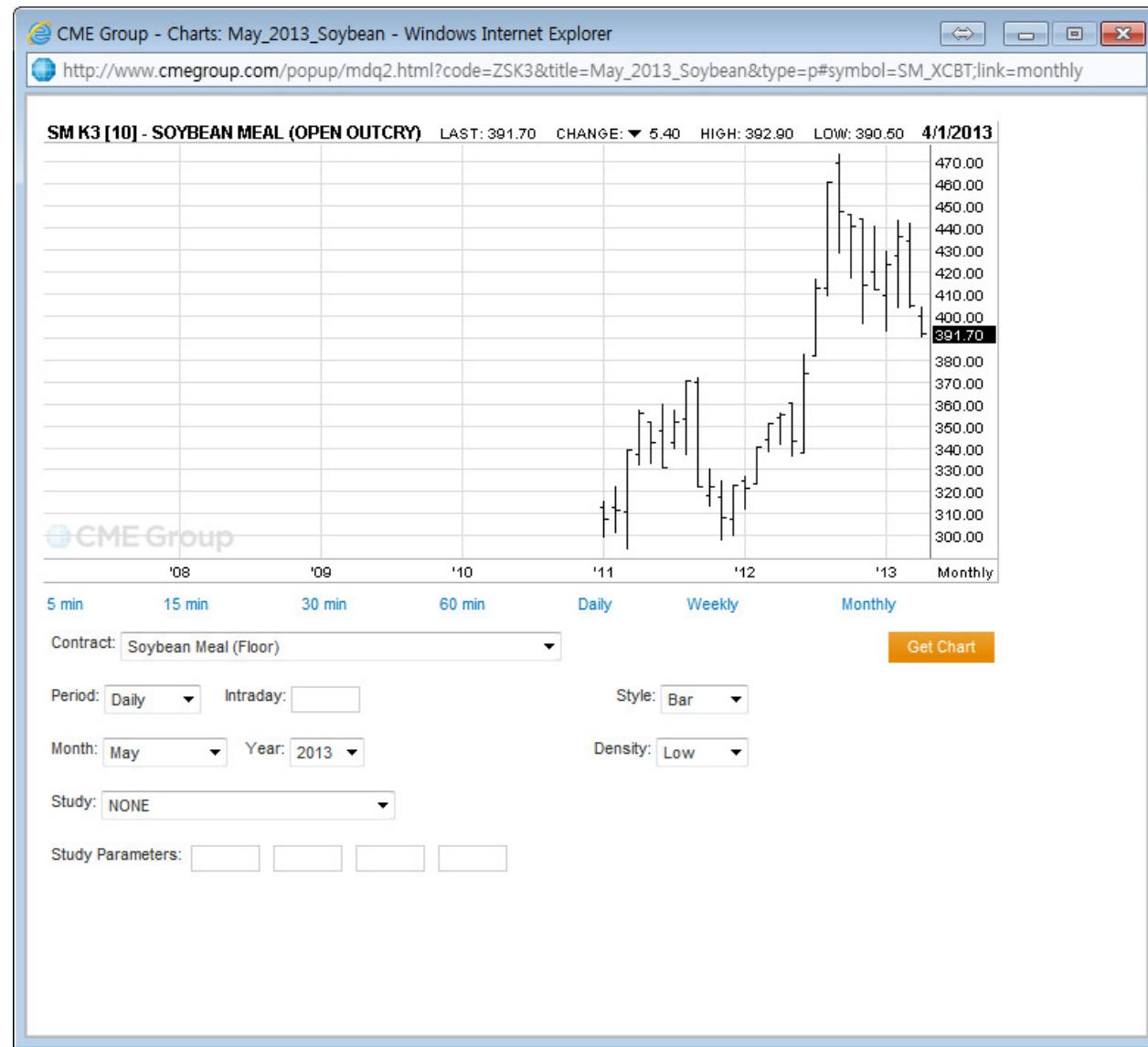


[http://www.geovar.com/data/satellite/rapideye/rapideye\\_sample\\_image\\_bands543.jpg](http://www.geovar.com/data/satellite/rapideye/rapideye_sample_image_bands543.jpg)

# Crop Yield Estimation



# Soybeans Futures (CME)



# Data Mining Methods



## *Decision Trees*

## *Naive Bayes*

## *Cluster Analysis*

# *Sequence Clustering*

## *Association Rules*

## *Time Series*

## *Neural Networks*

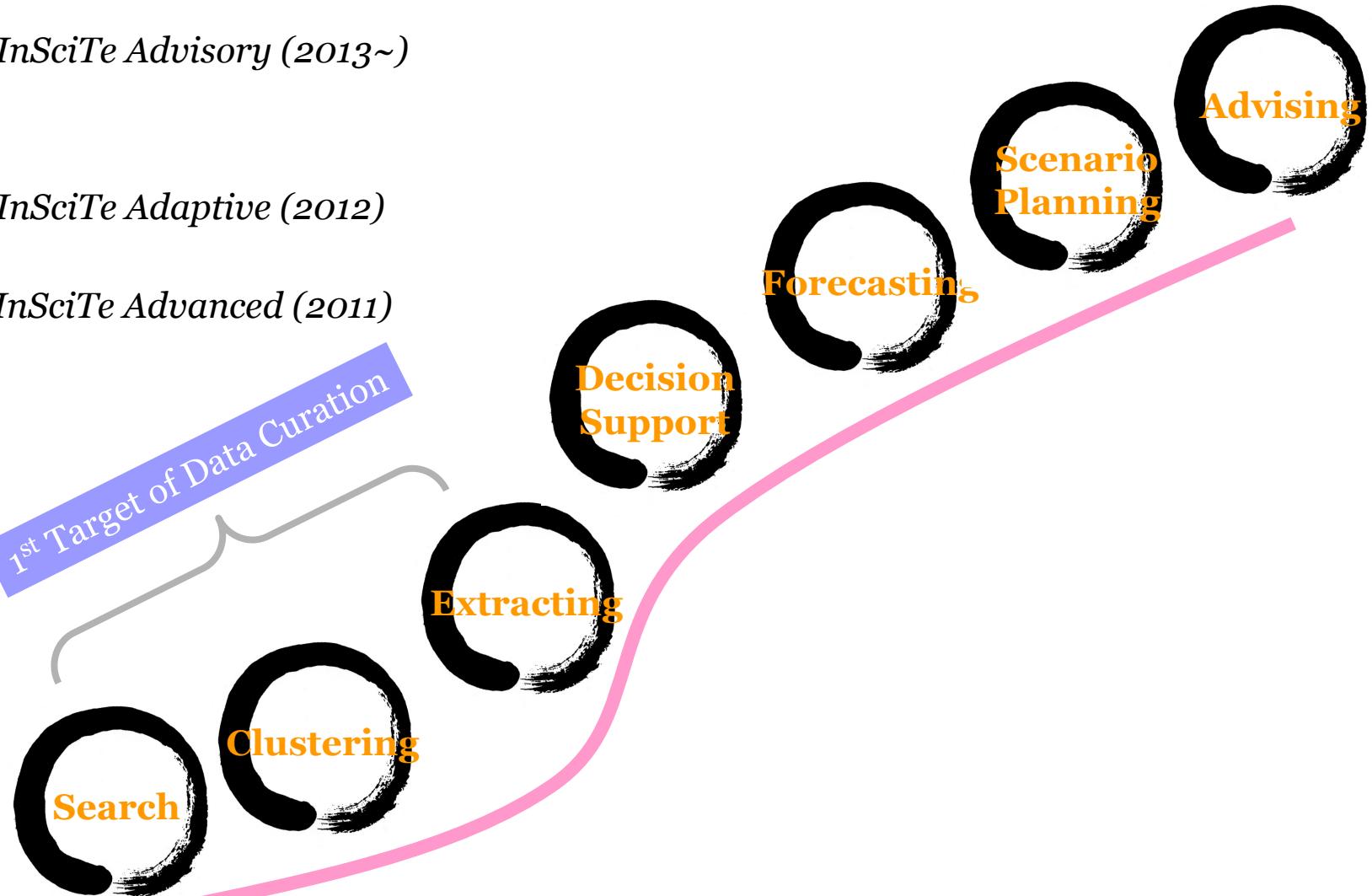
# Value Pyramid

*InSciTe Advisory (2013~)*

*InSciTe Adaptive (2012)*

*InSciTe Advanced (2011)*

1st Target of Data Curation



Modified from D. Bousfield & P. Fooladi, "STM Information: 2009 Final Market Size and Share Report", 2010.



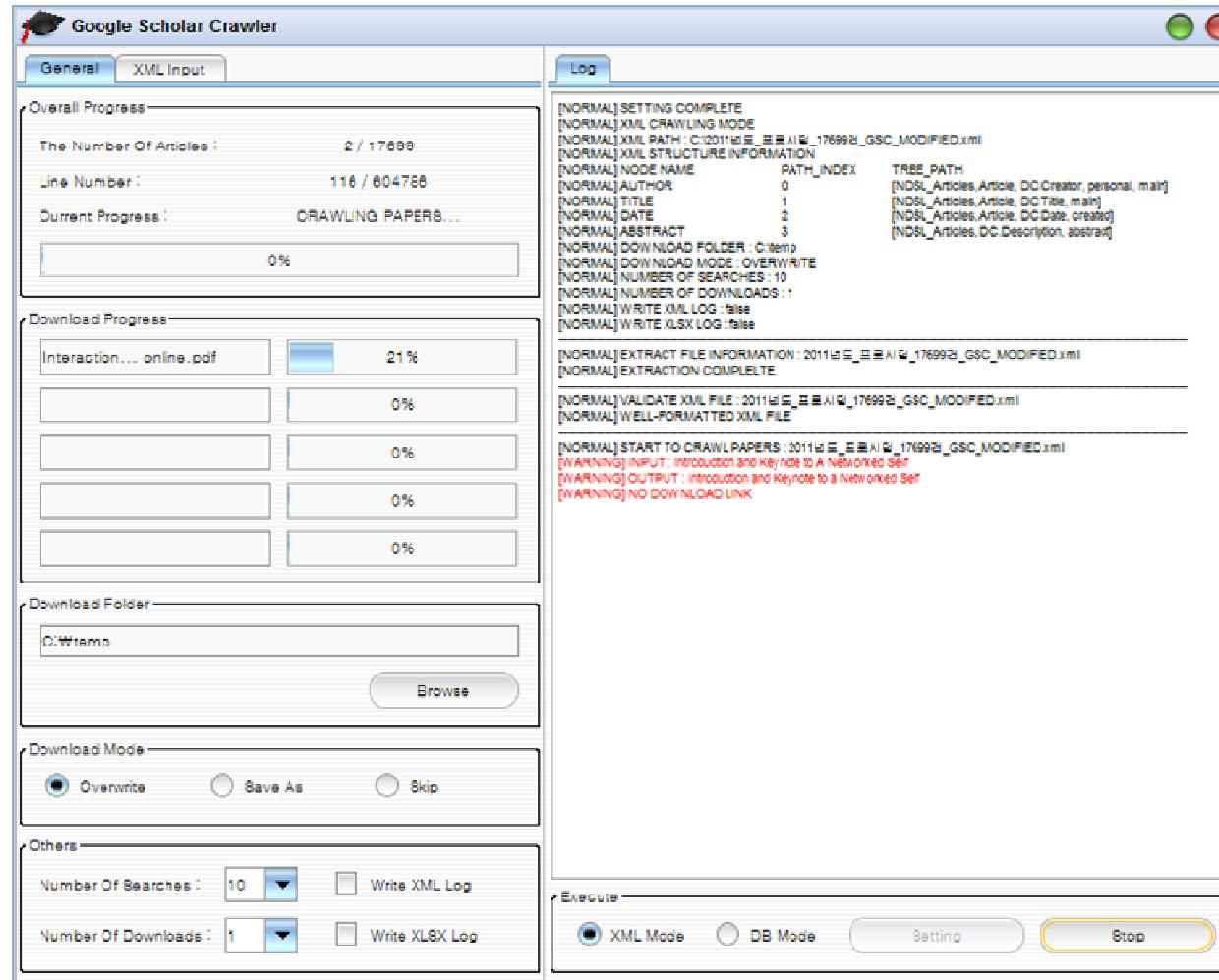
# Data Curation for Getting Insights & Foresights

---

- 1. Crawling Data*
- 2. Extracting Information (Text Mining)*
- 3. Resolving Identities*

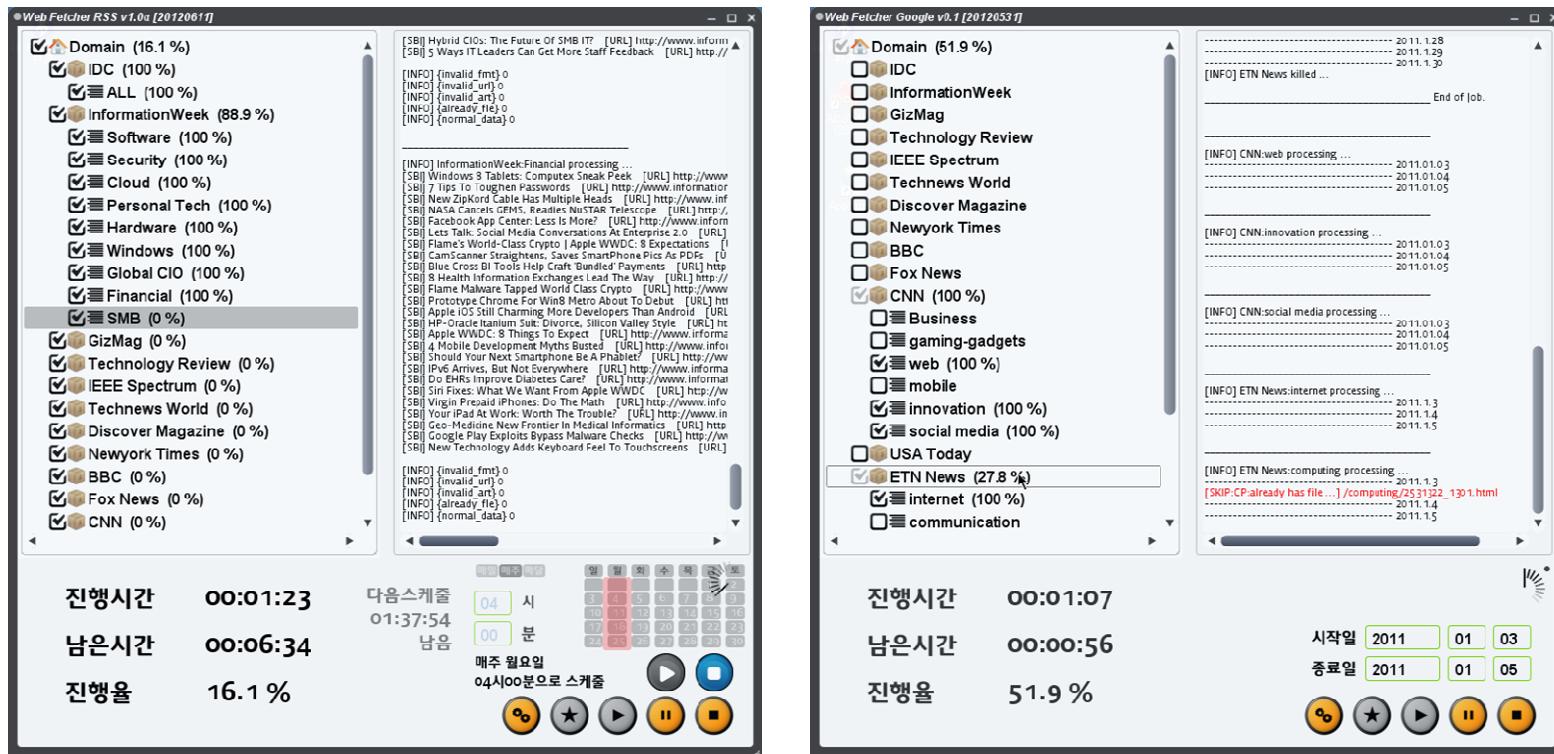
# Crawling Data

## ■ Crawling Full Texts from Google Scholar



# Crawling Data

## ■ Crawling Web Data by RSS & Google API

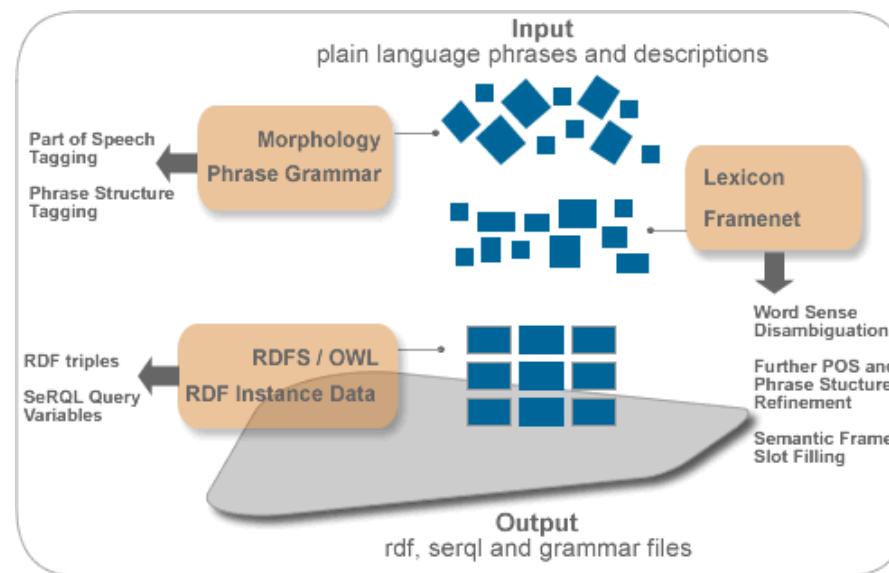


# Crawling Data

## ■ Web Data Statistics in InSciTe Adaptive

Source/Year	2001	...	2008	2009	2010	2011	2012	Total
IDC	0	...	1	13	21	313	250	598
Wikipedia	0	...	151,942	181,721	408,017	1,555,867	249,0209	4,975,178
InformationWeek	0	...	470	551	536	388	81	2,316
Gizmag	0	...	1,371	2,301	2,970	2,850	2,520	16,731
Technologyreview	89	...	646	693	758	873	405	5,330
IEEE spectrum	64	...	468	426	385	306	205	3,173
Technewsworld	31	...	696	742	1,396	2,332	2,485	9,843
DiscoverMagazine	2	...	104	51	79	57	49	606
NewYork Times	20,940	...	7,979	4,696	3,971	3,669	2,064	124,100
BBC	3,155	...	2,687	3,158	3,318	5,964	4,243	37,073
Fox News	0	...	1,577	1,965	1,123	2,015	1,872	10,749
CNN	3,594	...	1,154	1,499	1,071	1,308	814	19,769
Thomson Reuters	0	...	450	371	309	338	222	3,408
USA Today	458	...	4,616	3,878	4,630	6,579	4,276	38,750
EtnTws.com	447	...	1,964	2,241	2,099	1,585	844	14,259
Total	28,780	...	176,125	204,306	430,683	1,584,444	2,510,539	5,261,883

# Text Mining – Concept



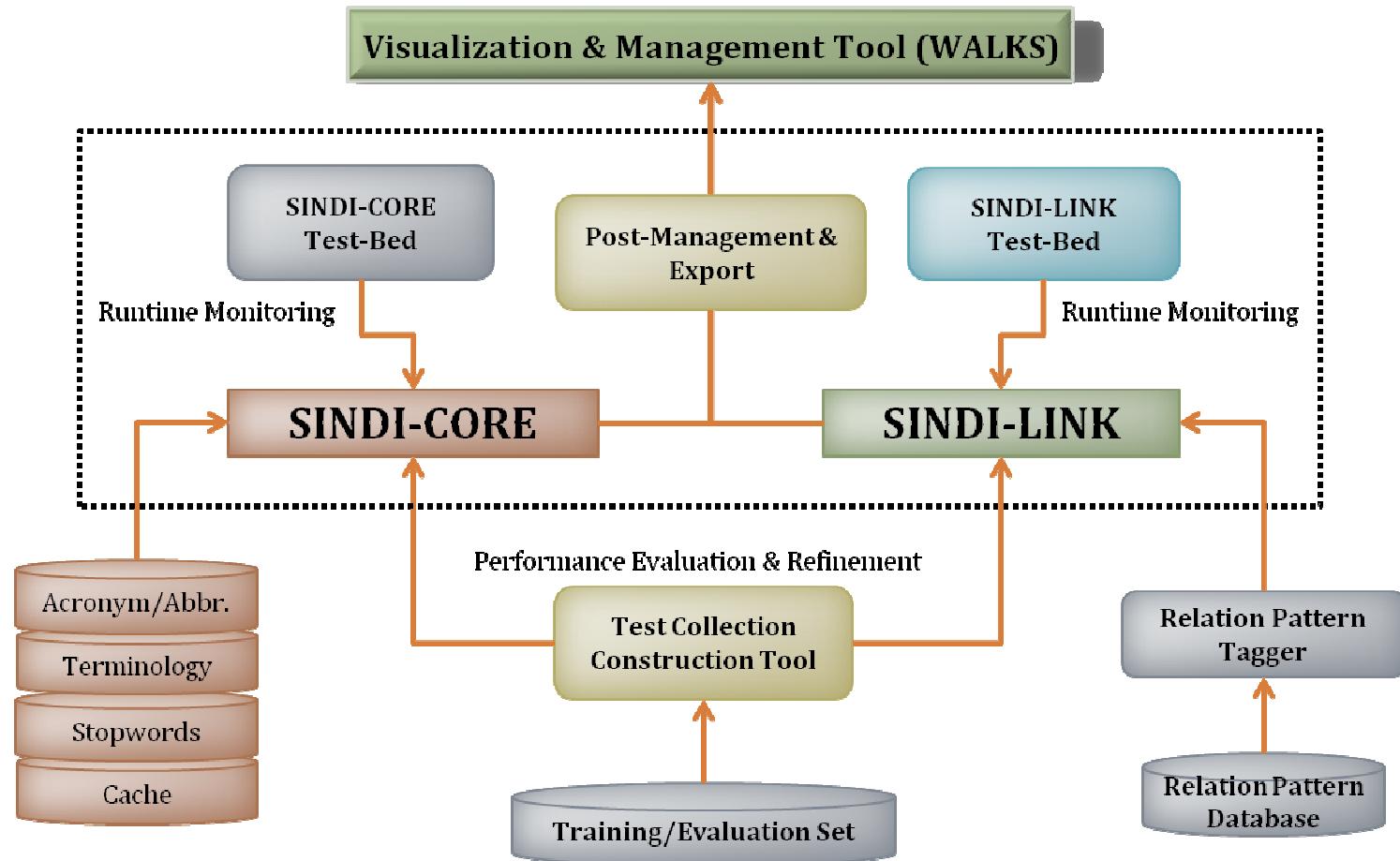
# Text Mining – Concept

---

## ■ Mining Targets

- Named entities: e.g. PLOs
- Pattern-based entities: e.g. e-mail addresses, phone numbers
- Concepts: abstractions of entities
- Facts and relationships
- Concrete and abstract attributes: e.g. 10-year, expensive, comfortable
- Subjectivity in the forms of opinions, sentiments, and emotions: e.g. positive/negative, angry

# Text Mining – SINDI Architecture



# Text Mining – Example

---

1 K1A1

[http://50.70.100.51:8080/Jane\\_Mines/JM\\_GBL001](http://50.70.100.51:8080/Jane_Mines/JM_GBL001)  
Janes\_Defence\_Equipment\_and\_Technology\janesxml\data\reference\jaa\jaa2011\jaa\_0043.xml

2 Leopard2

[http://50.70.100.51:8080/Jane\\_Mines/JM\\_GBL001](http://50.70.100.51:8080/Jane_Mines/JM_GBL001)  
Janes\_Defence\_Equipment\_and\_Technology\janesxml\data\reference\jaa\jaa2011\jaa\_0021.xml

3 M1A1

[http://50.70.100.51:8080/Jane\\_Mines/JM\\_GBL001](http://50.70.100.51:8080/Jane_Mines/JM_GBL001)  
Janes\_Defence\_Equipment\_and\_Technology\janesxml\data\reference\jaa\jaa2011\jaa\_0084.xml

4 Type90

[http://50.70.100.51:8080/Jane\\_Mines/JM\\_GBL001](http://50.70.100.51:8080/Jane_Mines/JM_GBL001)  
Janes\_Defence\_Equipment\_and\_Technology\janesxml\data\reference\jaa\jaa2011\jaa\_0039.xml

# Text Mining – Example

## Hyundai Rotem K1/K1A1 MBT

**Document Index:** Development, Description, K1A1 MBT, Armoured Vehicle-Launched Bridge, Armoured recovery vehicle, K1 mainclearing vehicle, Armoured engineer vehicle, K1-M MBT, K2 MBT, Specifications, Status, Contractor.

**Development**

Following proposals from a number of armoured fighting vehicle manufacturers, in 1980 the South Korean government selected the now General Dynamics Land Systems (at that time Chrysler) of the United States to design and build two prototypes of a new MBT to meet its own specific requirements.

The first of two prototypes of the XK-1 MBT was completed in 1983, the Automotive Test Rig (ATR). It was shipped to Aberdeen Proving Grounds in November 1983 for automotive performance, endurance and reliability testing. The ATR was a fully payloaded MBT chassis fitted with a non-operational turret.

The second prototype, called the Fire-Control Test Rig (FCTR), was rolled out at a ceremony at Selfridge Air National Guard base in December 1983 and shipped to Aberdeen Proving Grounds in February 1984 to begin fire control tests.

Production of the K1 began in South Korea in 1984 and first production vehicles were completed in 1985. The vehicle was subsequently type-classified as the K1 MBT and the tank made its first appearance in September 1987, by which time several battalions had been equipped with it. The first production batch of 210 vehicles was completed in 1987 with the second batch consisting of 325 vehicles.

Many key components of the K1, such as the General Dynamics Canada ballistic computer, MTU diesel engine, Renk transmission and Sagem roof-mounted sight were eventually manufactured in-country.

No production figures of the K1 (105 mm) and K1A1 (120 mm) MBTs have ever been released but it is estimated that at least 1,500 have been built for the Republic of Korea Army.

The vehicle has been offered on the export market as of early 2010 no export sales have been made. It is understood that production of the latest K1As is now complete. This will be followed in production by the K2 MBT which is covered in detail in a separate entry in Jane's Armour and Artillery. This features a crew of three and is armed with a 120 mm L/55 smooth bore gun.

**Description**

The layout of the K1 MBT is conventional, with the driver's compartment at the front, fighting compartment in the centre and engine and transmission at the rear.

Over the frontal arc of the K1 MBT, including the nose, hull front and sides, advanced armour of the Chobham type is fitted, which is manufactured in the US; this provides protection from both kinetic and chemical energy attack.

The driver is seated front left and is provided with a single-piece hatch cover pivoted on the left that lifts upwards to open. This has three integral day periscopes, the centre one of which can be replaced by a passive night driving periscope. The commander is seated on the right of the turret with the gunner below and forward of his position and the loader on the left.

The commander has a French Sagem (previously SFIM) two-axis independent stabilised 360° panoramic sight which has a magnification of x3 and x10, periscopes for all-round observation and a single-piece hatch cover that opens to the rear. This has been manufactured in South Korea by Samsung Electronics Co Ltd, under licence from Sagem (previously SFIM). This has three modes of operation; align with main armament, independent target tracking and gunner override.

Copyright © IHS Global Limited, 2010.

## Hyundai Rotem K1/K1A1 MBT

**Record Details:**

Title: Hyundai Rotem K1/K1A1 MBT  
Section: Tanks  
Country: Korea, South  
Organisation: ROTEM  
Images: 6  
Last posted: 2010-Mar-16  
Update status: Updated

**Publication Details:**

Publication: Jane's Armour and Artillery  
Edition: 2010

**Document Index:** Development, Description, K1A1 MBT, Armoured Vehicle-Launched Bridge, Armoured recovery vehicle, K1 mainclearing vehicle, Armoured engineer vehicle, K1-M MBT, K2 MBT, Specifications, Status, Contractor.

**Specifications**

Dimensions and weight	
<b>Crew:</b>	4
<b>Length</b>	9.672 m (31 ft 7 in)
<b>hull:</b>	7.477 m (24 ft 5 in)
<b>Width</b>	3.594 m (11 ft 9 in)
<b>Ground clearance</b>	0.46 m (1 ft 5 in)
<b>Weight</b>	51100 kg (112656.2 lb) (51 t)
<b>Ground pressure</b>	0.87 kg/cm² (8531 Pascals)
<b>Mobility</b>	
<b>Configuration</b>	tracked
<b>Powerto-weight ratio:</b>	23.48 hp/t
<b>Speed</b>	65 km/h (40.4 mph) 8 km/h (5 mph)
<b>Range</b>	500 km (310.7 miles)
<b>Ammunition:</b>	nn
<b>Fording</b>	1.2 m (3 ft 4 in) 2.20 m (7 ft 2 in)
<b>Gradient:</b>	60%
<b>Side slope:</b>	30%
<b>Vertical obstacle</b>	1.10 m (3 ft 6 in)
<b>Trench:</b>	2.74 m
<b>Engine</b>	MTU MB 871 Ka-501
<b>configuration:</b>	V-8

**Record Details:**

Title: Hyundai Rotem K1/K1A1 MBT  
Section: Tanks  
Country: Korea, South  
Organisation: ROTEM  
Images: 6  
Last posted: 2010-Mar-16  
Update status: Updated

**Publication Details:**

Publication: Jane's Armour and Artillery  
Edition: 2010

Copyright © IHS Global Limited, 2010.

# Text Mining – Example

The layout of the K1 MBT is conventional, with the driver's compartment at the front, fighting compartment in the centre and engine and transmission at the rear.

Context

Over the frontal arc of the K1 MBT, including the nose, turret front and sides, advanced armour of the Chobham type is fitted, which is manufactured in the US; this provides protection from both kinetic and chemical energy attack.

The driver is seated front left and is provided with a single-piece hatch cover pivoted on the left that lifts upwards to open. This has three integral day periscopes, the centre one of which can be replaced by a passive night driving periscope. The commander is seated on the right of the turret with the gunner below and forward of his position and the loader on the left.

Value

The commander has a French SAGEM (previously SFIM) two-axis independent stabilised 360° panoramic sight which has a magnification of  $\times 3$  and  $\times 10$  periscopes for all-round observation and a single-piece hatch cover that opens to the rear. This has been manufactured in South Korea by Samsung Electronics Co Ltd, under licence from SAGEM (previously SFIM). This has three modes of operation; align with main armament, independent target tracking and gunner override.

Weapon/Part

The gunner's two-axis stabilised day/night sight device incorporates a laser range-finder and thermal imaging system which is similar to that installed in the General Dynamics Land Systems M1A1 MBT (still used by the US Army and Marines) and has magnifications of  $\times 1$  and  $\times 10$  (day) and  $\times 3$  and  $\times 10$  (night). The gunner's articulated auxiliary sight is provided by the Electro-Optical Division of the Kollmorgen Corporation and Opto Mechanik Inc (OMI). This has a magnification of  $\times 8$ .

The latest Texas Instruments (now part of the Raytheon Systems Company) Gunner's Primary Tank Thermal Sight incorporates an eye-safe carbon dioxide laser range-finder with final integration and testing taking place in South Korea.

Turret drive and weapon elevation is electrohydraulic with manual controls for emergency use. Stabilisation is provided in both elevation and traverse.

The fire-control system includes a General Dynamics Canada digital ballistic computer with a number of sensors including crosswind and allows the tank to engage both stationary and moving targets while it is stationary or moving itself.

Main armament comprises a US-designed 105 mm M68A1 (which is based on the then UK Royal Ordnance Factories L7 rifled tank gun) rifled gun which is also fitted on many South

# Text Mining – Example

제원 분류		추출 대상	Specifications	Sentences
최상위 제원	상위 제원	상세 제원	제원 값	제원 값
Dimensions and weights (크기와 무게)	Crew (탑승인원)		4	
	Length (길이)	overall:	9.672 m (31.73 ft)	
		hull:	7.477 m (24.53 ft)	
	Width (폭)	overall:	3.594 m (11.79ft)	
	Ground clearance (지상고)	hull:	0.46 m (1.51ft)	
	Weight (무게)	combat:	51100 kg (112656.2 lb) (51.1t)	53.2 tones
Mobility (기동)	Configuration	running gear:	tracked	
		Power-to-weight ratio:	23.48 hp/t	
	Speed (속도)	top speed:	65 km/h (40.4 mph)	
		on 60 per cent gradient:	8 km/h (5 mph)	
	Range (기동거리)	main fuel supply:	500 km (310.7 miles)	
	Amphibious (수륙양용 여부)	amphibious:	no	
	Fording (도설 가능 깊이)	without preparation:	1.2 m (3.94 ft)	
		with preparation:	2.20 m (7.22 ft)	
	Gradient (경사도)	gradient::	60%	
	Side slope (비탈)	side slope:	30%	
	Vertical obstacle (수직 장애물)	forwards:	1.10 m (3.61 ft)	
	Trench (참호)	trench:	2.74 m	

# Text Mining – Example

최상위 제원	상위 제원	상세 제원	MBT			
			M1A2 (1992)	Leopard 2 (1979)	K1 (1987)	Type 90 (1991)
Dimensions and weights (크기와 무게)	Crew (탑승인원)		4	4	4	3
	Length (길이)	overall:	9.83 m (32.25 ft)	9.668 m (31.72 ft)	9.672 m (31.73 ft)	9.755 m (32 ft)
		hull:	7.92 m (25.98 ft)	7.722 m (25.33 ft)	7.477 m (24.53 ft)	7.5 m (24.61 ft)
		main armament rear:	9.033 m (29.64 ft)	8.49 m (27.85 ft)		
	Width (폭)	overall:	3.657 m (12 ft)	3.7 m (12.14 ft)	3.594 m (11.79 ft)	3.43 m (11.25 ft)
		without skirts:	3.479 m (11.41 ft)			3.33 m (10.93 ft)
	Height (높이)	to turret roof:	2.375 m (7.79 ft)	2.48 m (8.14 ft)		
		overall:				3.045 m (9.99 ft)
		hull:	1.727 m (5.67 ft)	1.769 m (5.8 ft)		2.335 m (7.66 ft)
		to top of roof mounted weapon:	2.885 m (9.47 ft)			
		axis of fire:	1.89 m (6.2 ft)	2.01 m (6.59 ft)		
	Ground clearance (지상고)	hull:	0.483 m (1.58 ft)		0.46 m (1.51 ft)	0.45 m (1.48 ft)
		hull sides:	0.432 m (1.42 ft)			
	Track (궤도)	front:		0.537 m (1.76 ft)		
		rear:		0.487 m (1.60 ft)		
		adjustable:				0.20 m (0.66 ft) to 0.60 m (1.97 ft)
	Track width (궤도 폭)	overall:		2.785 m (9.14 ft)		2.71 m (8.89 ft)
	Length of track on ground: (바닥접면궤도 길이)	normal:	635 mm	635 mm		600 mm
	Weight (무게)	combat:	63086 kg (139080.8 lb) (63.1 t)	55150 kg (121584.9 lb) (55.2 t)	51100 kg (112656.2 lb) (51.1 t)	50000 kg (110231.1 lb) (50 t)
	Ground pressure (접지압)	standard track:	1.08 kg/cm <sup>2</sup> (105912 Pa)	0.83 kg/cm <sup>2</sup> (81395 Pa)	0.87 kg/cm <sup>2</sup> (85318 Pa)	0.89 kg/cm <sup>2</sup> (87279 Pa)
	Configuration	running gear:	tracked	tracked	tracked	tracked
		Power-to-weight ratio:	23.77 hp/t	27.00 hp/t	23.48 hp/t	30.00 hp/t
	Speed (속도)	top speed:	67.6 km/h (42 mph)	72 km/h (44.7 mph)	65 km/h (40.4 mph)	70 km/h (43.5 mph)
		cross-country:	48.28 km/h (30 mph)			
		on 10 per cent gradient:	27.36 km/h (17 mph)			
		reverse:		31 km/h (19.3 mph)		42 km/h (26.1 mph)
		on 60 per cent gradient:	6.6 km/h (4.1 mph)		8 km/h (5 mph)	

# Text Mining – Application

This image shows a screenshot of a Korean defense procurement website. The top navigation bar includes the logo of the Defense Technology Agency (DTAQ) and the text "국방기술품질원" (National Defense Science & Technology Agency). The main search interface features tabs for "무기체계검색" (Weapon System Search) and "정보분석" (Information Analysis). A search bar at the top right contains the text "Smart Defense Leader". Below the search bar is a large search button labeled "전체" (All) with a checkmark icon.

The left sidebar displays a search result for the "40 mm L70/T92" system. It includes sections for "Title", "Statement", "Description", and "Development". The "Title" section lists "Taiwan develops new air-defense weapon" and "40 mm L70/T92". The "Statement" section notes that the system is designated as the 40 mm L70 T92 air-defense weapon and is due to go into service. The "Description" section provides a detailed technical description of the system's components and capabilities. The "Development" section mentions the Republic of China (Taiwan)'s development of a new towed low-level anti-aircraft gun (AAG) system based on a modified Swedish Bofors Defense 40 mm L/48 AAG system.

The central search interface allows users to search for weapon systems by category. The "무기체계" (Weapon System) dropdown menu includes categories like 국방정보체계, 감시/정찰무기체계, 지휘통제/통신체계, 지휘통제/통신체계, 험비무기체계, 방호무기체계, 항공무기체계, 험비무기체계, 기타무기체계, and 비무기체계. The "JDDS 분류체계" (JDDS Classification System) dropdown menu includes categories for Air and Land, such as Ammunition, Armored fighting vehicles, Explosive ordnance disposal, Homeland security equipment, Land vehicle propulsion, Land vehicle systems & equipment, Land weapon systems, Artillery, Infantry weapons, and Land-based air defence.

The search results for "40 mm L70/T92" show a detailed diagram of the system's internal components and their relationships. The diagram includes labels such as "Night operation", "Magazine", "Unit of fire", "Fording", "Feed", "Fuze", "Tyres", "Kill", "Elevation", "Cooler", "Warhead", "Recoil", "Track", "Crew", "Barrel life", "Fire", "Material", "Mount drive", "HEELbase", "Tracking", "Resolution", "Length", "Height", and "Specification".

On the right side of the search results, there is a "제원 속성 값" (Performance Parameters Value) section listing various specifications with checkmarks indicating they apply to the system. These include 20 mm close-in weapon system, anti-aircraft heavy machine gun, Nd:YAG with Raman-shift, towed anti-aircraft gun (AAG), twin towed anti-aircraft gun, and two section blast continuous-rod fragmentation with piercing effect.

At the bottom right, there is a "Copyright © 2013 Hanmin Jung" watermark.

# Resolving Identities

## ■ URI Aliases

- URIs that refer to the same real-world objects
  - E.g. <http://dbpedia.org/resource/Berlin> (for Berlin in DBpedia)
  - E.g. <http://sws.geonames.org/2950159> (for Berlin in Geonames)
- Information providers can set owl:sameAs links to URI aliases they know about

## ■ Resolution of Data Conflicts in Data Fusion

- Choosing a value in situations where multiple sources provide different values for the same property of an object



Christian Bizer, Tom Heath, and Tim Berners-Lee, "Linked Data – The Story So Far", 2009.

# Resolving Identities

---

## ■ Example

- 소녀시대, 소시, 少女時代, しょうじょじだい, 少時, *Sho Jo Ji Dai, So Nyeo Shi Dae, SoShi, Girl's Generation, SNSD, So Nyuh Shi Dae, ...*



# Resolving Identities

The screenshot shows the Sindice search interface. At the top, there is a logo for "indice THE SEMANTIC WEB INDEX" and a navigation bar with links for About, Search, Submit Your Data, Jobs, Support, and Dev. A status message "web01 Version: 2.1.16" is also present. Below the header, the search interface type is set to "Simple". The search term "Hanmin Jung" is entered in the keyword(s) field. The search results are displayed under the heading "Sindice search: Hanmin Jung found 116 documents in 0.19 seconds". Each result entry includes the document type (e.g., RDF, XFN, SPARQL), the date it was created, the number of triples, and the size. It also provides a link to the document and options to search or inspect it.

Search interface type: [Simple](#) [Advanced](#) [Guru](#) [Query Language Documentation](#)

keyword(s)

**SEARCH** Group By Dataset:  Sorted by: relevance ▾

Quick filters ([All options](#))

Time range: [Any date](#)  
[Today](#) [Yesterday](#) [Last week](#)  
[Last month](#) [Last year](#)

Format: [Any format](#)  
[RDF](#) [RDFa](#) [MICRODATA](#)  
[MICROFORMAT](#) [XFN](#) [HCARD](#)  
[HCALENDAR](#) [LISTING](#)  
[HRESUME](#) [LICENSE](#) [GEO](#)  
[ADR](#)

Predicate:

Class:

Ontology:

Domain:

Sindice search: Hanmin Jung found 116 documents in 0.19 seconds

**Hanmin Jung** (RDF)  
+ 2010-07-19 – 27 triples in 9.1 kB  
<http://data.semanticweb.org/person/hanmin-jung> ([Search](#)) [Inspect](#) ([Cache](#)) ([Live](#))

**"Hanmin Jung"** (SPLITTER\_FORMAT)  
+ 2012-06-18 – 5 triples in 655 Bytes  
[http://dblp.i3s.de/d2r/resource/authors/Hanmin\\_Jung](http://dblp.i3s.de/d2r/resource/authors/Hanmin_Jung) ([Search](#)) [Inspect](#) ([Cache](#)) ([Live](#))

**Hanmin Jung | ResearchGATE** (RDFA)  
+ 2011-01-08 – 3 triples in 384 Bytes  
[http://www.researchgate.net/profile/Hanmin\\_Jung](http://www.researchgate.net/profile/Hanmin_Jung) ([Search](#)) [Inspect](#) ([Cache](#)) ([Live](#))

**Hanmin Jung** (RDF)  
+ 2008-12-12 – 3 triples in 603 Bytes  
<http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff4753885839df-4204c51622a647afa10cac9b6228a...> ([Search](#)) [Inspect](#) ([Cache](#)) ([Live](#))

**Hanmin Jung** (RDF)  
+ 2008-12-16 – 4 triples in 806 Bytes  
<http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff4753885839df-46bdf8ab26a5d49b67335df767060...> ([Search](#)) [Inspect](#) ([Cache](#)) ([Live](#))

**Hanmin Jung** (RDF)  
+ 2008-12-16 – 3 triples in 597 Bytes  
<http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff4753885839df-870ee64b05206fd12c924d3b53fa...> ([Search](#)) [Inspect](#) ([Cache](#)) ([Live](#))

**Hanmin Jung** (RDF)  
+ 2008-12-12 – 3 triples in 594 Bytes  
<http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff4753885839df-7c10e325448048842168929bc7dd3...> ([Search](#)) [Inspect](#) ([Cache](#)) ([Live](#))

**Hanmin Jung** (RDF)  
+ 2008-12-16 – 3 triples in 594 Bytes

<http://sindice.com/search?q=Hanmin+Jung>

# OntoURIResolver

## Demonstration

Powered by 

OntoURIResolver

Keyword URI Hanmin Jung  
Examples : Hanmin Jung, Alok N. Choudhary, Compilation for Distributed Memory

keyword : Hanmin Jung

- Summary Information

**Hanmin Jung , person ( 9 URIs )**

- Resolved URI List

Canonical Group : Hanmin Jung [9] click

URI : (<http://data.semanticweb.org/person/hanmin-jung...>)

- [1. http://acm.rkbexplorer.com/id/person-666054-d3d227a15e3...](http://acm.rkbexplorer.com/id/person-666054-d3d227a15e3...)
- [2. http://acm.rkbexplorer.com/id/person-575114-63a9445ee7d...](http://acm.rkbexplorer.com/id/person-575114-63a9445ee7d...)
- [3. http://citeseer.rkbexplorer.com/id/resource-CSP152100-4...](http://citeseer.rkbexplorer.com/id/resource-CSP152100-4...)

Show 5 more

- [9. http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...](http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...)

- UnResolved URI List

Deprecated URIs [14]

- [1. http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff47...](http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff47...)
- [2. http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff47...](http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff47...)
- [3. http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff47...](http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff47...)

Show 10 more

- [14. http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff47...](http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c3ff47...)

Non-Rdf URIs [2]

- [1. http://ontoworld.org/wiki/Special:URIResolver/Hanmin\\_Jung](http://ontoworld.org/wiki/Special:URIResolver/Hanmin_Jung)
- [2. http://www.researchgate.net/profile/Hanmin\\_Jung](http://www.researchgate.net/profile/Hanmin_Jung)

- Results from sameAs.org

1.Hanmin Jung [17]  
(<http://acm.rkbexplorer.com/id/person-575114-63a9445ee7d...>)

- [1. http://acm.rkbexplorer.com/id/person-575114-63a9445ee7d...](http://acm.rkbexplorer.com/id/person-575114-63a9445ee7d...)
- [2. http://acm.rkbexplorer.com/id/person-575114-ca70b05cd60...](http://acm.rkbexplorer.com/id/person-575114-ca70b05cd60...)
- [3. http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...](http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...)

Show 13 more

- [17. http://kisti.rkbexplorer.com/id/PER\\_0000000000000005510...](http://kisti.rkbexplorer.com/id/PER_0000000000000005510...)

2.Hanmin Jung [4]  
(<http://dblp.rkbexplorer.com/id/people-ca099511bbbb...>)

- [1. http://acm.rkbexplorer.com/id/person-666054-d3d227a15e3...](http://acm.rkbexplorer.com/id/person-666054-d3d227a15e3...)
- [2. http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...](http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...)
- [3. http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...](http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...)
- [4. http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...](http://dblp.rkbexplorer.com/id/people-ca099511bbbb21b9c...)

3.Hanmin Jung [2]  
(<http://data.semanticweb.org/person/hanmin-jung>)

- [1. http://data.semanticweb.org/person/hanmin-jung](http://data.semanticweb.org/person/hanmin-jung)
- [2. http://ontoworld.org/wiki/Special:URIResolver/Hanmin\\_Jung](http://ontoworld.org/wiki/Special:URIResolver/Hanmin_Jung)

4.Hanmin Jung | ResearchGATE [1]  
([http://www.researchgate.net/profile/Hanmin\\_Jung](http://www.researchgate.net/profile/Hanmin_Jung))

- [1. http://www.researchgate.net/profile/Hanmin\\_Jung](http://www.researchgate.net/profile/Hanmin_Jung)

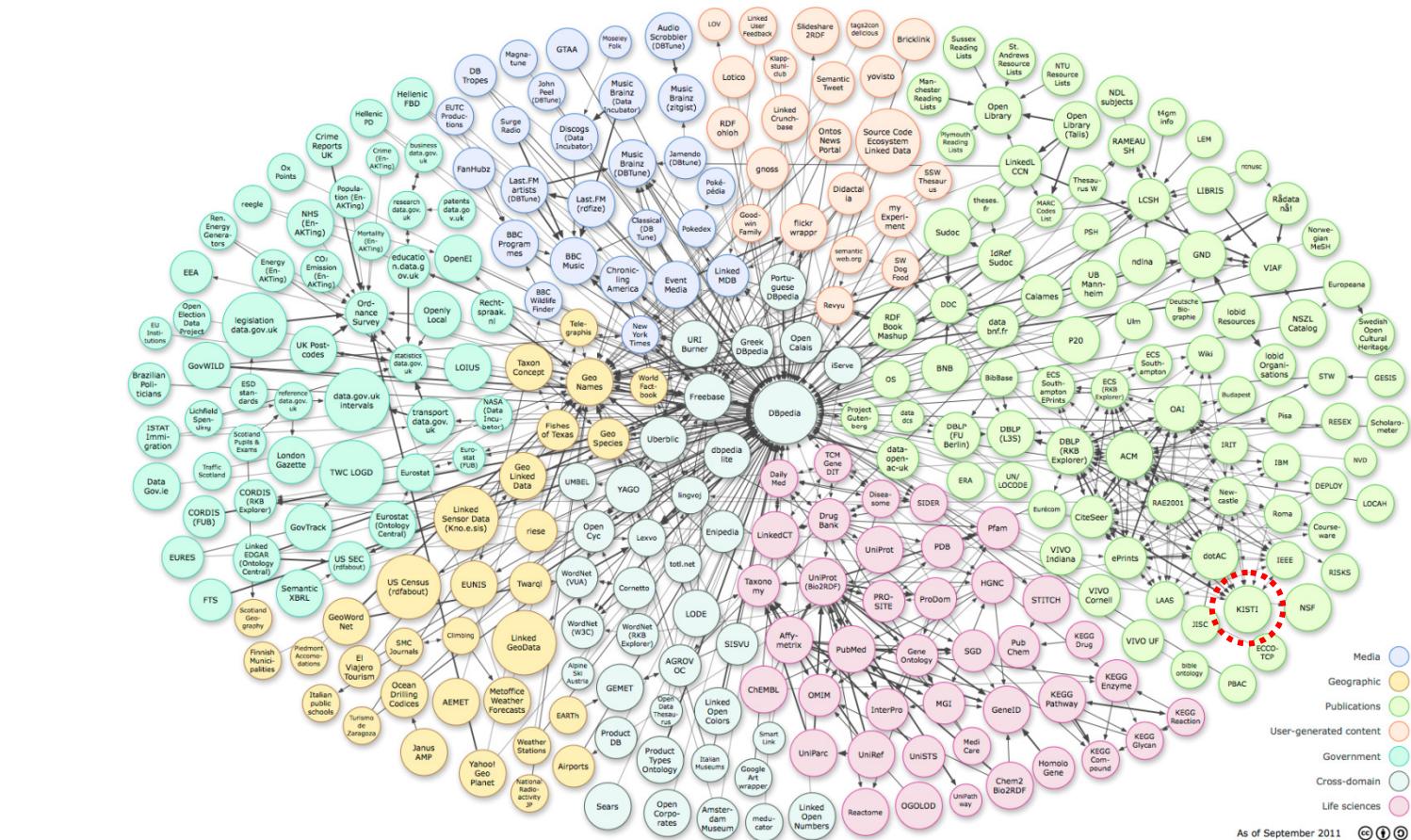
5.Hanmin Jung [1]  
([http://dblp.I3s.de/d2r/resource/authors/Hanmin\\_Jung...](http://dblp.I3s.de/d2r/resource/authors/Hanmin_Jung...))

- [1. http://dblp.I3s.de/d2r/resource/authors/Hanmin\\_Jung](http://dblp.I3s.de/d2r/resource/authors/Hanmin_Jung)

# LOD Project

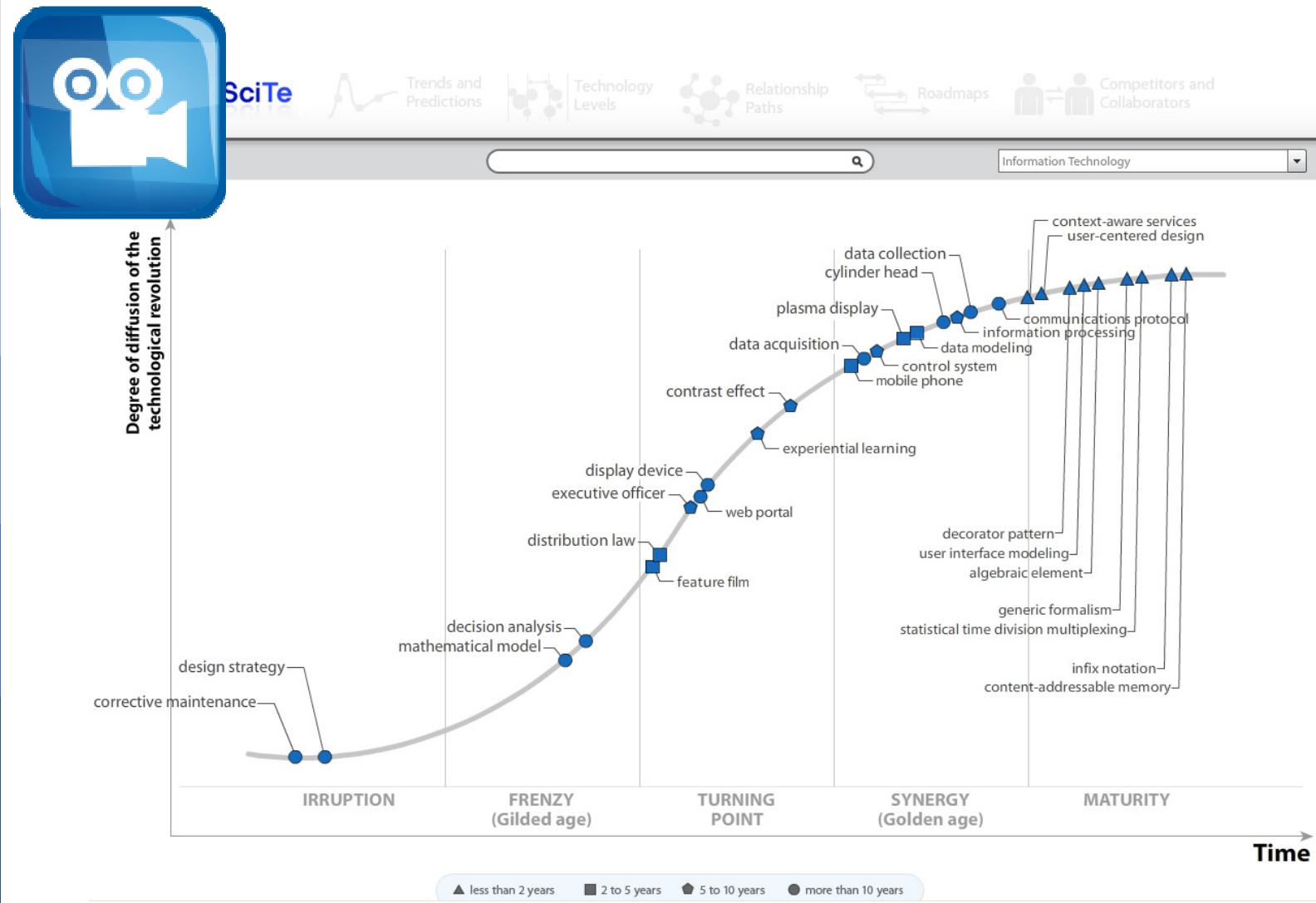
## ■ Linked Data

- 31 billion RDF triples, 504 RDF links (2011.9)

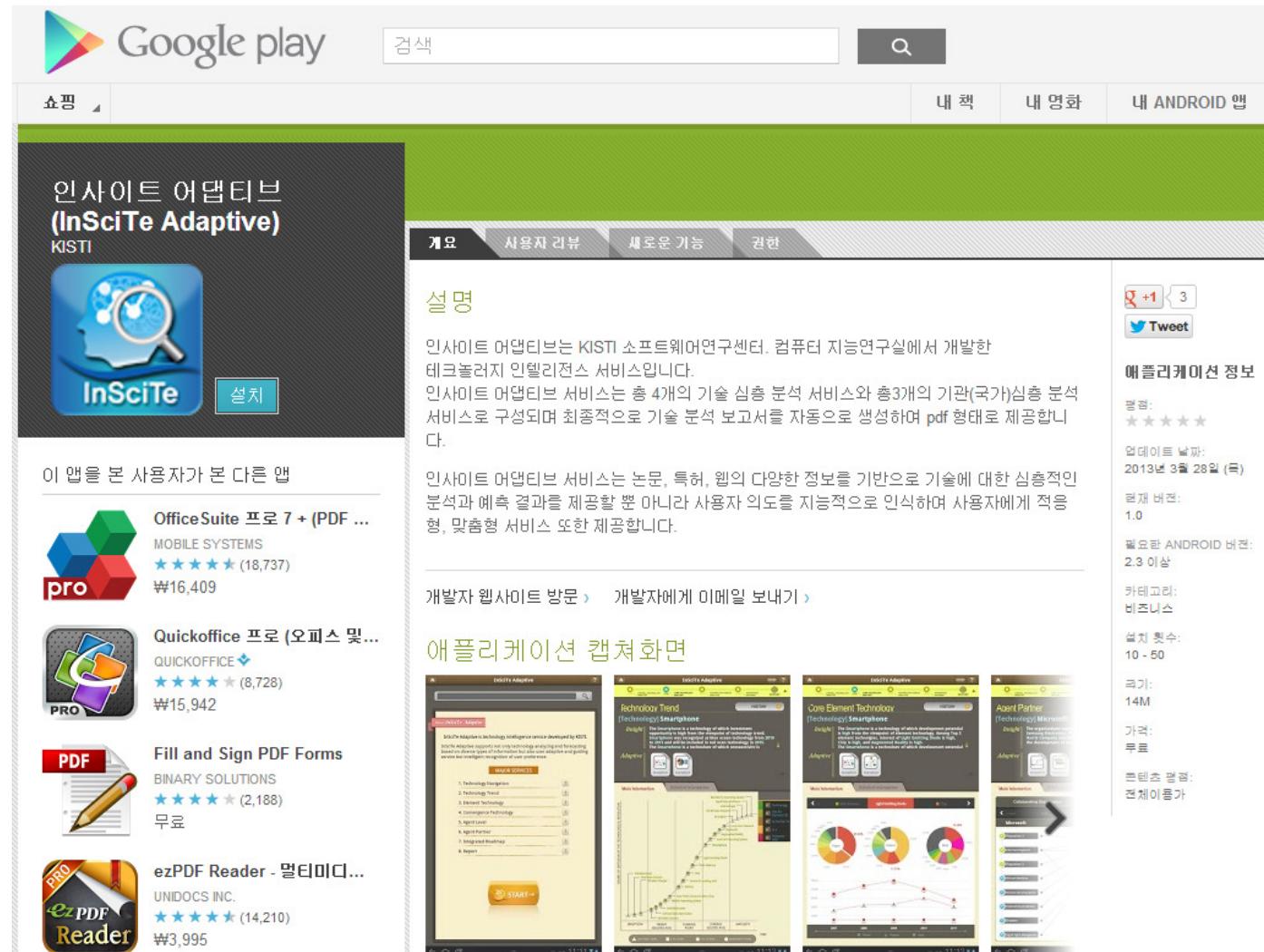


[http://richard.cyganiak.de/2007/10/lod/lod-datasets\\_2011-09-19\\_colored.html](http://richard.cyganiak.de/2007/10/lod/lod-datasets_2011-09-19_colored.html)

# InSciTe Advanced (2011)



# InSciTe Adaptive (2012)



[https://play.google.com/store/apps/details?id=net.xenix.inscite&feature=search\\_result#?t=W10](https://play.google.com/store/apps/details?id=net.xenix.inscite&feature=search_result#?t=W10).

# InSciTe Adaptive (2012)



## InSciTe Adaptive Report

### Augmented Reality

<u>1. Introduction</u>	<u>3</u>
1-1 Definition	4
1-2 Technology Overview	4
<u>2. Technology In-Depth Analysis</u>	<u>6</u>
2-1 Technology Trends	7
2-2 Core Elementary Technology	8
2-3 Convergence Technology	9
<u>3. Organization / Nation In-Depth Analysis</u>	<u>11</u>
3-1 Organization Level	12
3-2 Nation Level	13
3-3 Organiztion Partners	14
<u>4. Roadmap</u>	<u>15</u>
<u>5. Conclusion</u>	<u>18</u>

# InSciTe Adaptive (2012)

---

## ■ Data Fact Sheet

- Articles: 22.6 millions (9.8 millions for papers, 7.6 millions for patents, 5.3 millions for Web data)
  - All technical areas (2001~2011)
- Named entities: 1.9 millions
- Authority dictionary: 1.5 millions entries
- LOD data: 290 GB (*are being connected*)

## 맞춤형 ‘정보추출 연구’ 앱으로 구현

### KISTI ‘인사이트 어댑티브’… 엔진성능도 네 배 높여

한국과학기술정보연구원(KISTI)의 정보추출 연구가 스마트폰 앱으로 구현됐다.

KISTI 컴퓨터지능연구실은 인사이트 어댑티브(InSciTe Adaptive) 앱을 개발해 최근 구글 플레이(Google Play)에 등록했다고 16일 밝혔다.

인사이트 어댑티브는 연구개발 전략수립지원 서비스 앱이다.

네 개의 기술 심층 분석 서비스와 세 개의 기관(국가)심층 분석 서비스로 구성됐다.

기술 분석 보고서를 자동으로 생

성, PDF 파일 형태로 제공한다. 논문과 특히, 웹의 다양한 정보에 기반을 둔 기술의 심층적인 분석과 예측 결과를 제공할 뿐만 아니라 사용자의 도를 지능적으로 인식해 사용자 맞춤형으로 제공된다.

이번에 적용된 정보추출엔진은 기존 KISTI가 보유하고 있던 정보추출엔진 대비 네 배 이상 성능이 향상됐다. 또 2200만건의 빅데이터 문헌 데이터에 적용됐다.

이 앱은 과학기술 전문 글로벌 출판사 인테크(InTech)에서 출간된 ‘고급 텍스트 마이닝 이론과 응용’

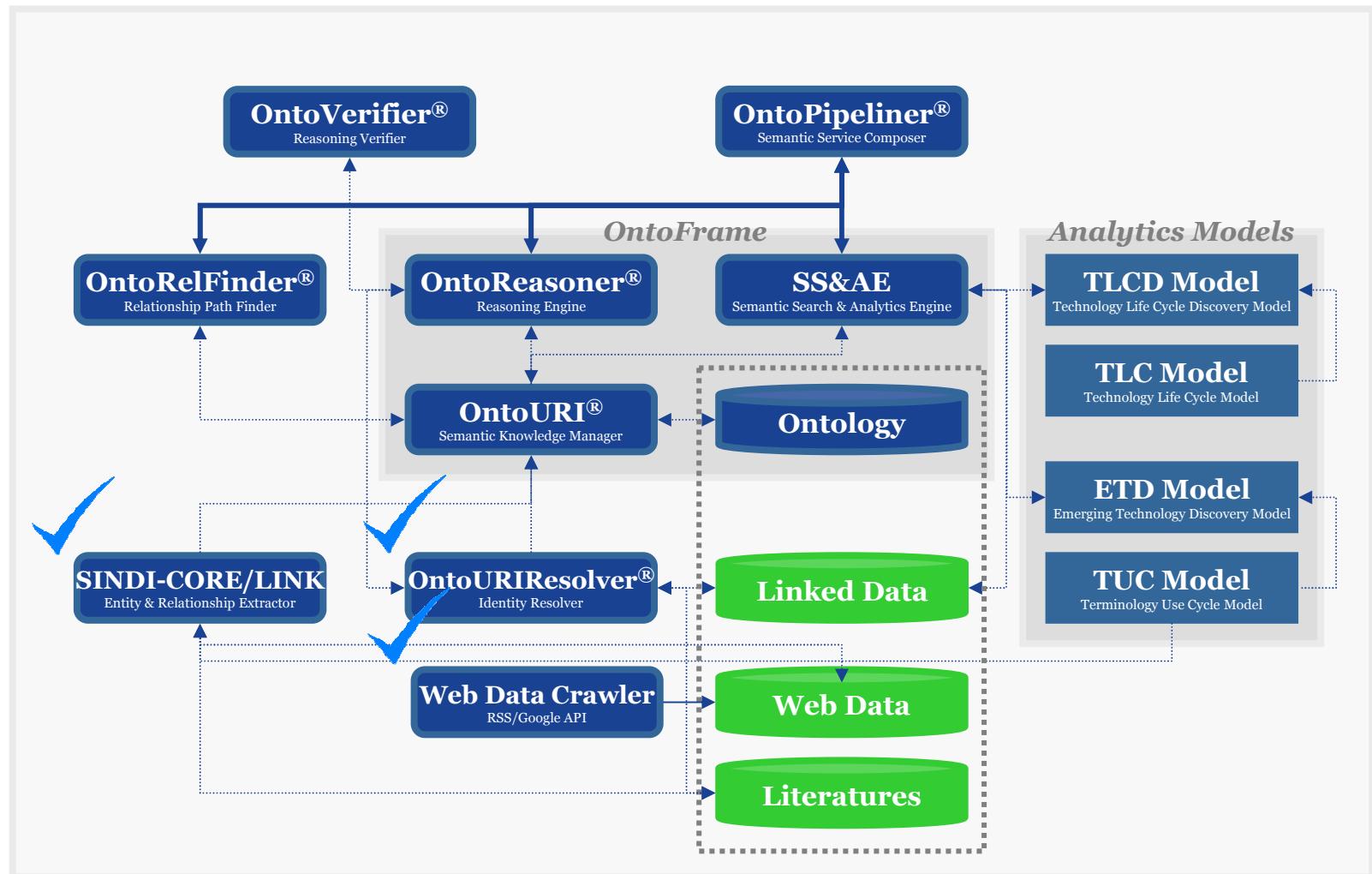
에 실린 KISTI의 ‘커널 기반 정보추출 조사’를 토대로 만들어졌다.

세계 20여명의 연구자가 참여해 KISTI가 작성한 챕터가 4개월 만에 다운로드 1000회를 돌파할 정도로 해당 도서 내에서 가장 많은 다운로드 수를 기록하며 세계 연구자의 주목을 받고 있다.

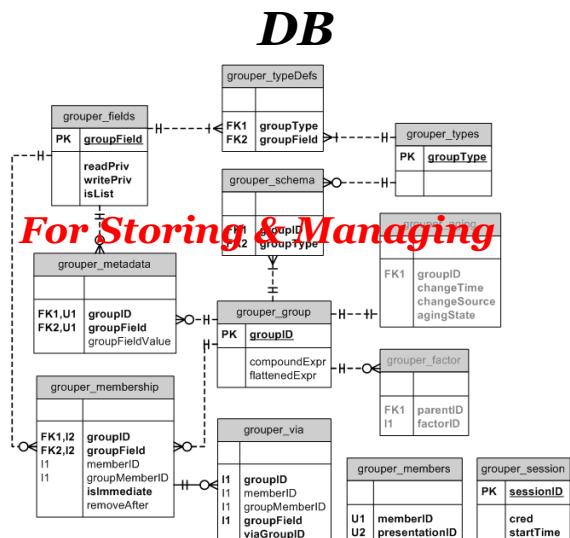
정한민 KISTI 컴퓨터지능연구실장은 “인사이트를 이용해 연구 개발 시간과 비용을 줄여 연구자뿐만 아니라 신규 사업 아이템 발굴에 고충을 겪는 중소기업에도 많은 도움을 줄 것으로 기대된다”고 밝혔다.

류경동기자 ninano@etnews.com

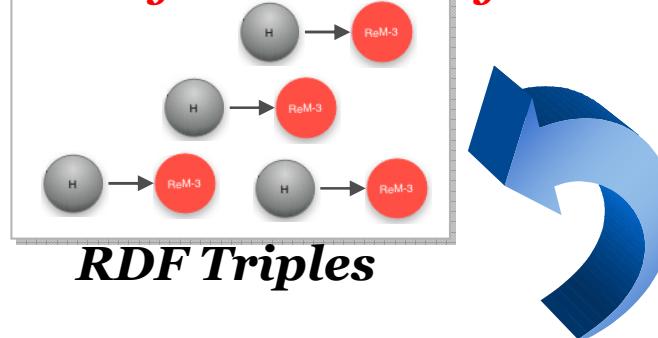
# InSciTe Architecture



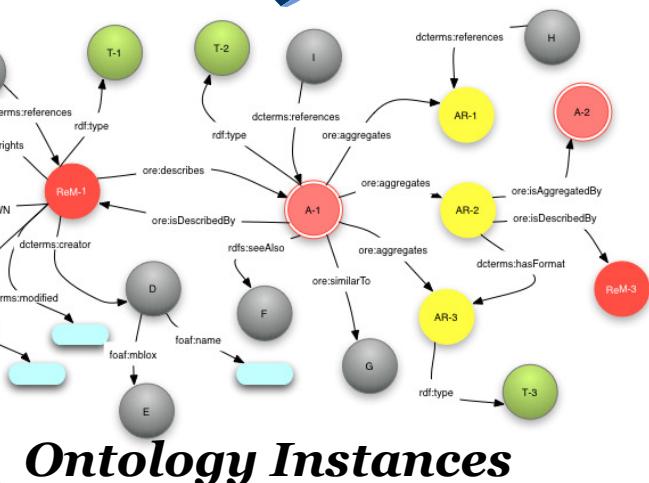
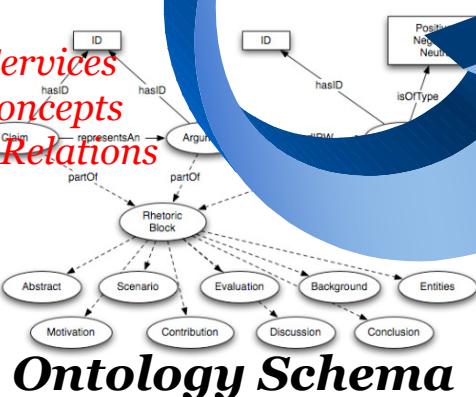
# Ontology



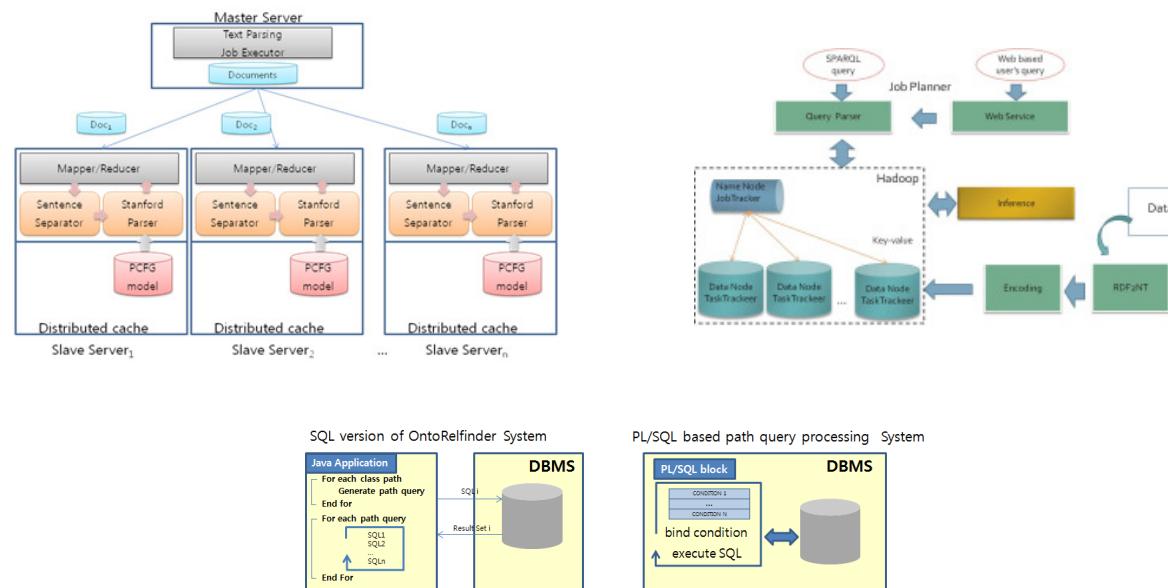
*Portability & Connectibility*



**For Service**  
*Planning Services  
Defining Concepts  
Exploiting Relations*



# Big Data Processing in InSciTe



# InSciTe Homepage

<http://inscите.kisti.re.kr>

InSciTe: Intelligence in Science & Technology					
InSciTe (Recent Version)	KISTI@Wikipedia	CFP@KISTI	Members	Publications	Case Studies
	<b>Version</b>	Main Functions	Development Duration	Documentation (TSCC Codec for Movie)	
	<b>InSciTe Adaptive</b>	Project in progress	Jan 01, 2012 - Dec 31, 2012	Project in progress	<b>Movie</b> <ul style="list-style-type: none"><li>- InSciTe Advanced (Eng.)</li><li>- InSciTe Advanced (Kor.)</li></ul> <b>Presentation</b> <ul style="list-style-type: none"><li>- InSciTe Advanced (Eng.)</li><li>- InSciTe Advanced Snapshot Scenario (Kor.)</li><li>- Narration</li><li>- For Movie</li><li>- InSciTe Overview Specification (Kor.)</li></ul>
	<b>InSciTe Advanced</b>	<b>Intelligent Decision Support Service</b> <ul style="list-style-type: none"><li>- Technology Trends</li><li>- Emerging Technologies</li><li>- Technology Levels</li><li>- Relationship paths</li><li>- Development Roadmaps</li><li>- Competitors and Collaborations</li></ul>	Jan 01, 2011 - Dec 31, 2011	<b>Movie</b> <ul style="list-style-type: none"><li>- InSciTe Advanced (Eng.)</li><li>- InSciTe Advanced (Kor.)</li></ul> <b>Presentation</b> <ul style="list-style-type: none"><li>- InSciTe Advanced (Eng.)</li><li>- InSciTe Advanced Snapshot Scenario (Kor.)</li><li>- Narration</li><li>- For Movie</li><li>- InSciTe Overview Specification (Kor.)</li></ul>	
	<b>InSciTe</b>	<b>Technology Intelligence Service</b> <ul style="list-style-type: none"><li>- Semantic Text Mining Techniques</li><li>- Report Services of Technologies and Organizations</li><li>- Connections with LOD</li></ul>	Jan 01, 2010 - Dec 31, 2010	<b>Movie</b> <ul style="list-style-type: none"><li>- InSciTe Presentation</li><li>- InSciTe 2010</li><li>- Intelligent Systems Snapshot</li><li>- InSciTe Brochure (HWP) Scenario (Kor.) Specification (Kor.)</li></ul>	<b>Movie</b>



“A lot of times, people don’t know what they want until you show it to them.”

*by Steve Jobs*

“Many people won’t be convinced until they’ve seen it for themselves.”

*by Jakob Nielsen*

*Thank you*

*jhm@kisti.re.kr*