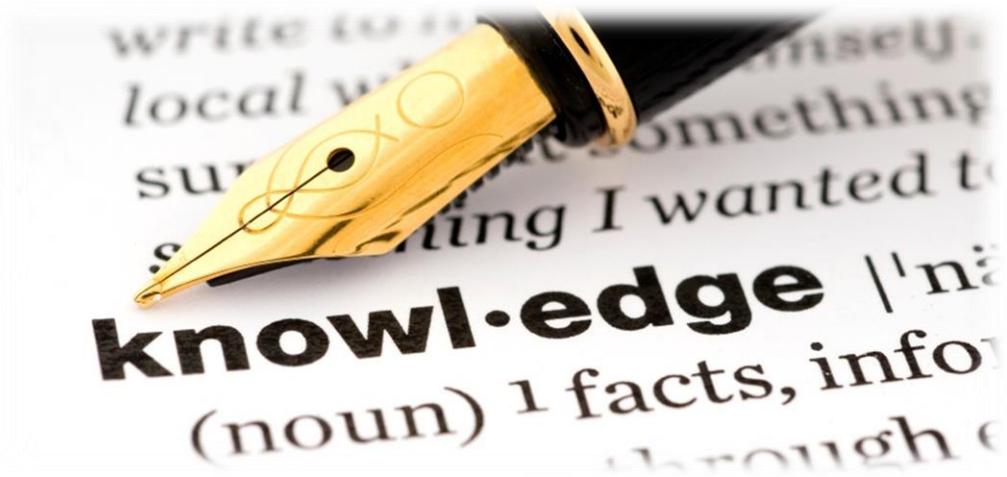


뉴스기사 지식화 및 분석 서비스

- 동향, 기본 개념 및 추진 방향 -



Prof. Sung-Pil Choi

Department of Library and Information Science

Kyonggi University

sungpil@gmail.com, spchoi@kgu.ac.kr

2015-05-07

발표순서

- ▶ 서론 (“危機” 시리즈)
- ▶ 국내외 뉴스기사 지식 서비스 사례
- ▶ 뉴스기사 지식베이스(**News Article Knowledge Base**)
- ▶ 뉴스기사 분석서비스(**News Article Analysis Service**)
- ▶ 추진 전략 및 체계
- ▶ 기대 효과
- ▶ 결론

Can Superman Save Daily Planet?



James Curran, "The Future of Journalism"

저널리즘의 위기



신문사의 위기

Philadelphia Inquirer Newsroom

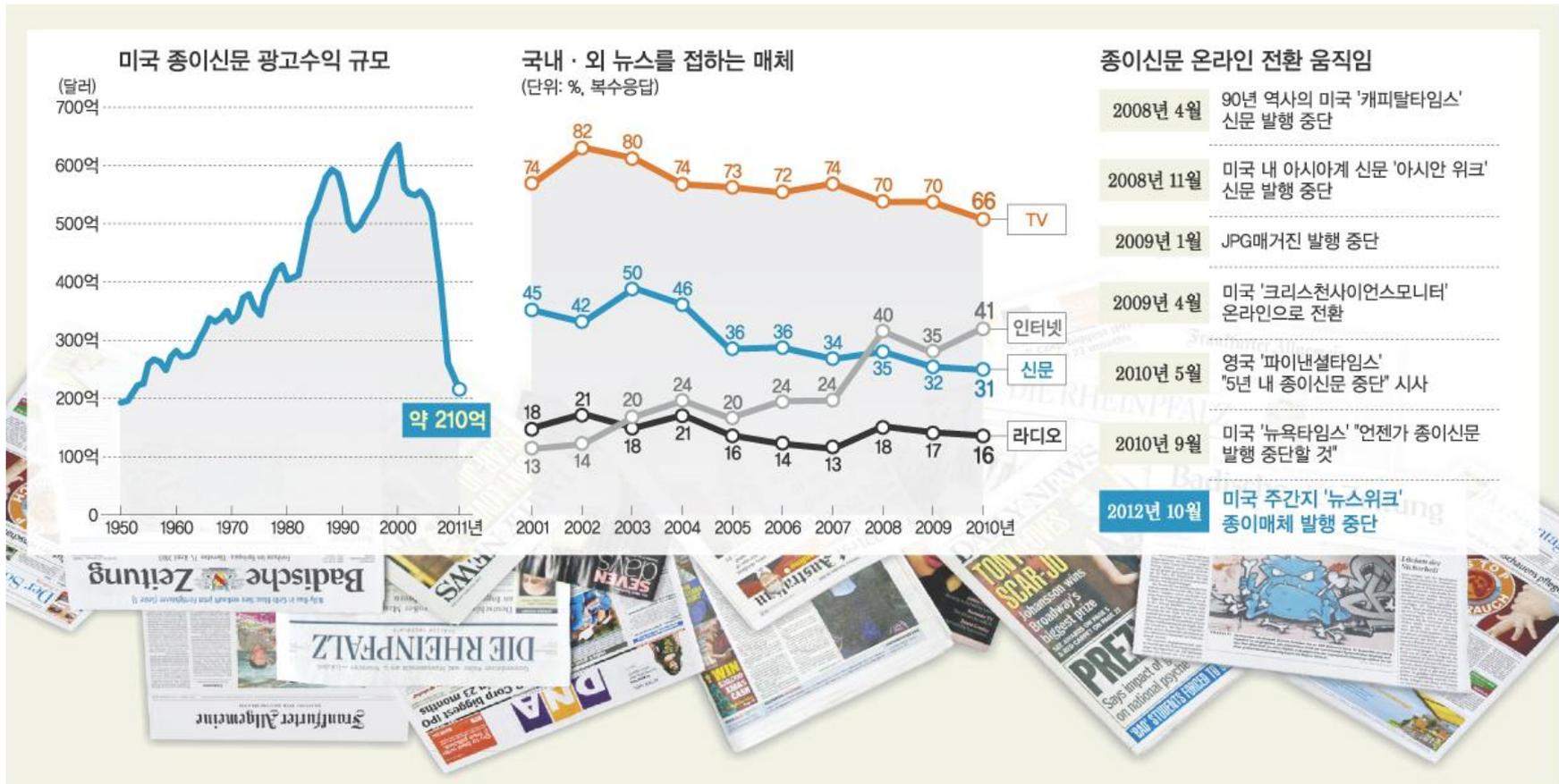


2009



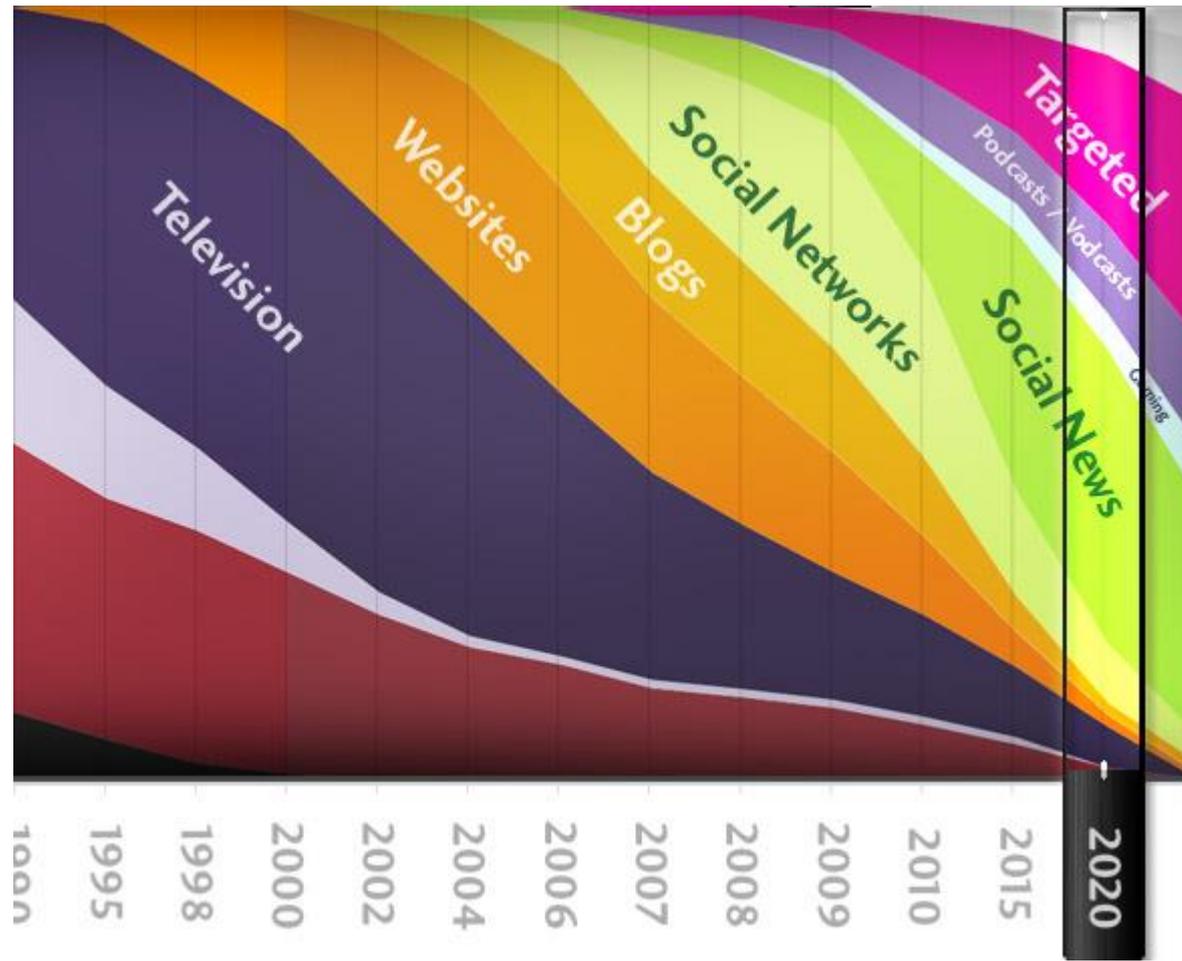
2012

신문사의 위기



<http://m.segye.com/view/20121021021761>

신문사의 위기



국내 신문사의 위기

홈 > 미디어뉴스 > 뉴스

사라진 '포털 어드밴티지', '뉴스의 위기' 보여주는 조선닷컴 트래픽 1등신문 닷컴도 휘청거리는 상황 "한국 온라인저널리즘에는 관계도 기술도 없다"

2014.12.31 09:21:32

박장준 기자 weshe@mediaus.co.kr

한국 언론의 위기 '생계형 언론'



손 정 언
전 전남일보 편집국장

지난 1991년 9월 김중배 동아일보 편집국장이 사주의 편집국 개입에 맞서다 해임됐다. 이는 시장 권력과 자본 권력이 언론을 잠식한 대표적인 사건이었고 이후 '기자의 사원화'로 생계형 언론이 출현하게 되면서 한국 언론의 위기를 초래했다.

생계형 언론의 문제점은 '공론의 장'으로서의 사명보다는 자사 이익의 확대

재생산과 언론 권력을 추구하는데 있다. 이에 대한 보도행태는 왜곡된 의제 설정, 정파적 보도, 특정 이익을 위한 물타기 보도, 혼계성 보도 등으로 나타난다.

생계형 언론과 관련된 이슈는 최근 '성추행 의혹'으로 해임된 윤창중 대변인의 사례에서 찾아 볼 수 있다. 그는 기사를 통해 생계형 권력주의를 추구했다. 또 공정보도에 기여한 언론인으로 평가 받는 손석희의 종합편성채널 JTBC행은 찬반 논란을 부추겼다.

생계형 언론의 대안으로 떠오르고 있는 예는 기자가 주주로 참여해 협동조합으로 전환한 프레시안을 꼽을 수 있다. 이는 언론의 사명과 정체성 강화에 생존 길을 모색하고 있는 것으로 해석된다.

뉴스의 미래 (1) : 저널리즘 시스템의 위기 [특집]

필자: 강정수 | 작성일: 2012-03-26 | 카테고리: 미디어, 비즈니스 | 댓글: 4

“언론의 위기? 왜 망하는지조차 모르고 있다”

[인터뷰] 윤지영 오가닉미디어랩 대표, “미디어 플랫폼의 붕괴, 진화하지 않으면 죽는다”

입력 : 2014-03-09 20:51:34 | 노출 : 2014.03.10 11:42:39

이정환 기자 | black@mediatoday.co.kr

[신문의 날 특집] 위기의 종이신문, 운명의 기로에서 새로운 길을 모색할 때

김민섭 기자 | minseob2580@yonsei.ac.kr

사진으로 기록한 종이신문의 위기

허핑턴포스트코리아 | 작성자 허완

게시일: 2014년 08월 25일 17시 11분 KST | 업데이트됨: 2014년 08월 25일 17시 16분 KST



제57회 '신문의 날' 맞았지만...“신문의 위기 심각”

언론노조-민주당, 신문산업진흥특별법 입법 촉구

김순택 기자 | nanan@mediaus.co.kr

“신문산업 위기 인터넷 때문 아니다”

AUTHOR PMKMK2013 | PUBLISHED SEPTEMBER 30, 2013 | COMMENTS 0

설성인 조선비즈 기자 seol@chosun.com | 박민경 조선비즈 인턴기자 techchosun@chosun.com

한국 신문위기의 원인

: 뉴스 매체의 기능적 대체, 저가치 제공, 그리고 공정성 위기

이준웅**

(서울대학교 언론정보학과)

최영재***

(한림대학교 언론정보학부)

국내 신문사의 위기



위기의 원인

▶ 뉴스 미디어의 기능적 도태 및 대체

- ▶ 신규성, 신속성, 상호 작용성 부족
- ▶ 불확실 정보의 비교/검증 기능 부재
- ▶ 다양한 관점의 제공, 고급 논평, 심층적 분석 부족

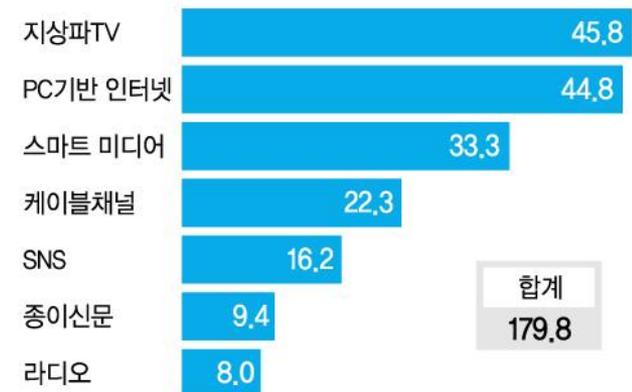
▶ 뉴스 미디어의 가치 저하

- ▶ 뉴스 미디어의 품질 및 서비스 정체
- ▶ 단편적인 단위 미디어에 대한 효용성 부족

▶ 뉴스 미디어의 공정성, 경영능력 부족

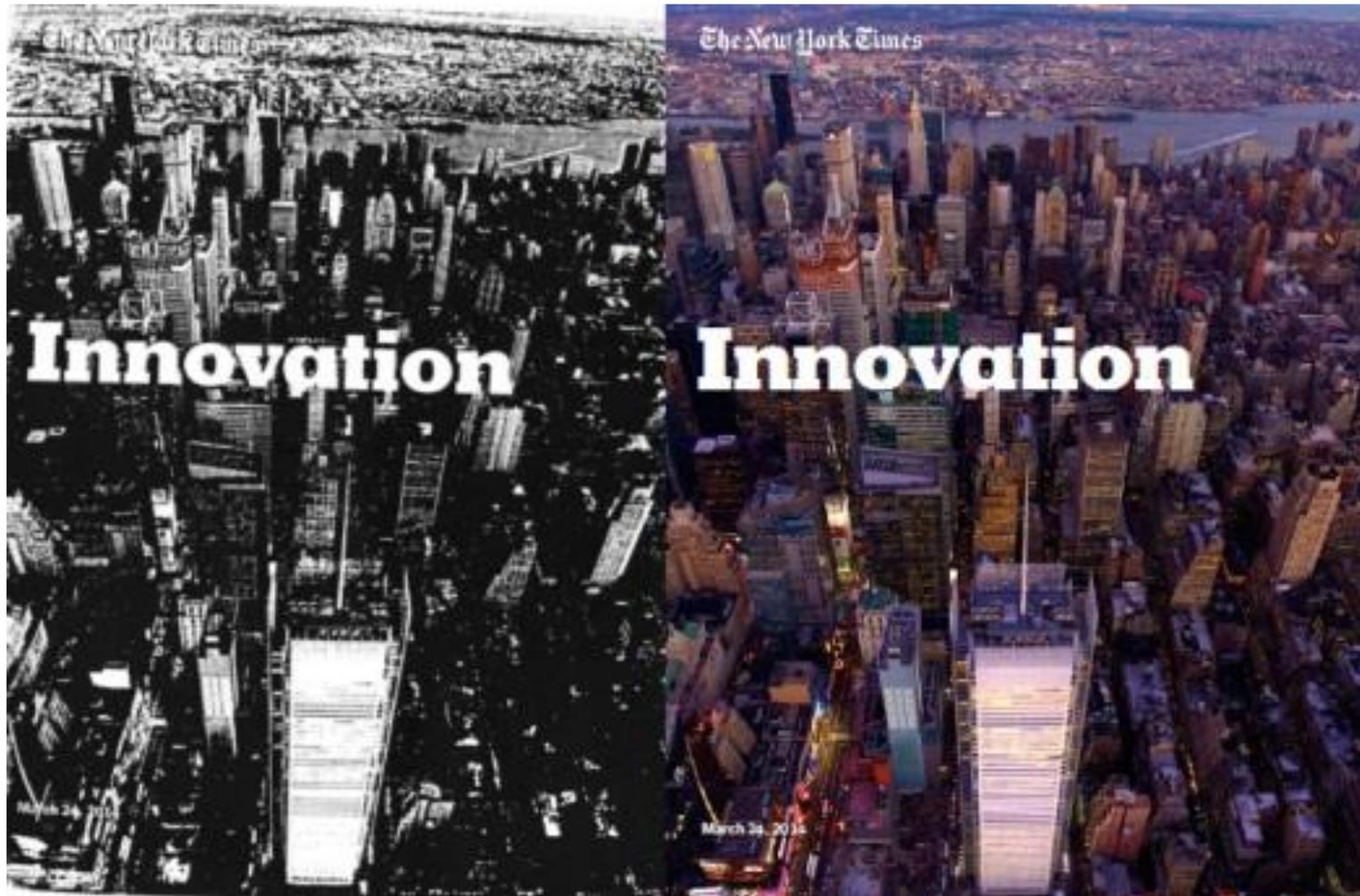
- ▶ 불공정성, 정치적 편파성
- ▶ 전사적, 공격적 경영체제 수립 미비

7개 플랫폼별 하루 평균 뉴스 이용량
(단위: 분, 20세 이상 1079명 설문조사)



자료: 한국언론진흥재단, '스마트 미디어 뉴스 생태계의 혁신전략'(2014)

위기를 기회로...



위기를 기회로...

▶ 뉴욕타임즈 혁신보고서

- ▶ 2014년 5월 15일 온라인 뉴스 및 엔터테인먼트 서비스 회사인 버즈피드(BuzzFeed), 매셔블에 의해서 공개
- ▶ 6개월간 354명(내부: 228명, 외부: 126명)과의 인터뷰를 통해 작성됨 (97페이지)
- ▶ 뉴스 소비 행태의 변화로 미디어 업계가 빠르게 재편되는 환경을 진단하고, 뉴욕타임스가 '디지털 퍼스트'를 추구해야 한다고 조언
- ▶ 전 세계에서 디지털 혁신에 가장 앞장서고 있는 언론사인 뉴욕타임즈의 자기 반성

위기를 기회로...

▶ 뉴욕타임즈 혁신보고서의 주요 내용 (강정수 2014)

혁신 요구 사항	내용
뉴스 도달거리 확장 방안 마련	<ul style="list-style-type: none"> · 온라인 뉴스기사 서비스에 적합한 구조화된 저널리즘 (Structured Journalism) 체제 구축 · 소셜 미디어 기반 서비스 체제 강화 · 뉴스 소비의 개인화 지원
편집국과 비즈니스 팀 사이의 협력	<ul style="list-style-type: none"> · 독자 및 청중 분석 데이터, 뉴스 유통 관련 통계 정보 등의 상호 공유
편집국에 전략팀 배치	<ul style="list-style-type: none"> · 내/외부 정보 수집, 기술 혁신 흐름 분석 등을 통한 새로운 프로젝트 제안 및 실행이 가능한 전략팀을 편집국 내에 배치하여 상시 운영
디지털 우선 전략(Digital First) 수립	<ul style="list-style-type: none"> · 종이신문의 제약에서 벗어나 디지털 뉴스 생산에 가능한 모든 역량 투여 · 최종적으로 디지털 뉴스 중에서 가장 훌륭한 뉴스를 선별해 다음 날 (아침) 종이신문에 배치하는 체제 구축

위기를 기회로...



*Digital = Content
First = First*

SearchLove San Diego 2014

@adammonago

구조적 저널리즘(Structured Journalism)

▶ 구조적 저널리즘 – Story Stream & Card Stacks

VIEW THE **NEWEST FIRST** **OLDEST FIRST** **MOST COMMENTED**

UPDATE 97
5 months ago
251

Apple files another request to ban older Samsung phones and tablets

By **Jacob Kastrenakes** on December 27, 2013 12:11 pm

Apple is once again seeking a sales ban on over 20 different Samsung smartphones and tablets that were found to violate its patents last year. Though such a ban was initially rejected, last month Apple **was given another chance** to argue for one when an appellate court vacated the earlier ruling. Apple filed its renewed request for a permanent injunction yesterday, arguing that Samsung's ability to sell the infringing devices puts a potential burden on its smaller lineup of smartphones and...

[Continue reading >](#)

UPDATE 96
6 months ago
82

Apple paid its lawyers over \$60 million to beat Samsung in court

By **Bryan Bishop** on December 6, 2013 02:27 pm

The damages retrial for last year's Apple vs. Samsung legal battle recently wrapped, and new court documents reveal how much Cupertino paid its legal team to get its wins: over \$60 million. The revelations come as Apple requests that Samsung reimburse a portion of those legal costs. **According to recent court filings**, Apple wants its rival to pay a third of that, coming out to more than \$15.7 million in attorney's fees — and that's on top of the over \$6.2 million in trial-related costs Samsung...

[Continue reading >](#)

Vox SHARE CARD STACK f t g

CARD 1 OF 17

Everything you need to know about network neutrality

BY TIMOTHY B. LEE
MAY 21, 2014, 2:53P

1. **What is network neutrality?**
2. Who invented the concept of network neutrality?
3. What's the argument for network neutrality?
4. Is regulation necessary to protect network neutrality?
5. What's the case against network neutrality regulations?
6. How did the government promote open networks in the internet's early days?

What is network neutrality?

Consumers generally **connect to the internet** one of two ways. They can subscribe to a residential broadband service from a company such as Time Warner Cable. Or they can subscribe to **wireless internet access** from companies such as Sprint.

These companies have spent billions of dollars laying cables in the ground (in the case of residential internet access) or erecting cell phone towers (for wireless access) to ensure that customers have fast, reliable service.

Network neutrality is the idea that these companies should treat all internet traffic equally. It says your ISP shouldn't be allowed to block or degrade access to certain websites or services, nor should it be allowed to set aside a "fast lane" that allows content favored by the ISP to load more quickly than the rest.

Since the term was coined more than a decade ago, it has been at the center of the debate over internet regulation. **Congress**, the **Federal Communications Commission (FCC)**, and the **courts** have all debated whether and how to protect network neutrality.

Advocates argue that network neutrality lowers barriers to entry online, allowing entrepreneurs to create new companies like Google, Facebook, and Dropbox. But **critics warn** that regulating the broadband market could be counterproductive, discouraging investment in internet infrastructure and limiting the flexibility of ISPs themselves to innovate.

In January, an appeals court **invalidated FCC regulations** designed to protect network neutrality. The agency is currently **considering how to respond**.

SHARE THIS CARD f t g

버지(theVerge.com)의 삼성과 애플의 경쟁에 관한 이야기 흐름

복스(Vox.com)의 카드형 뉴스

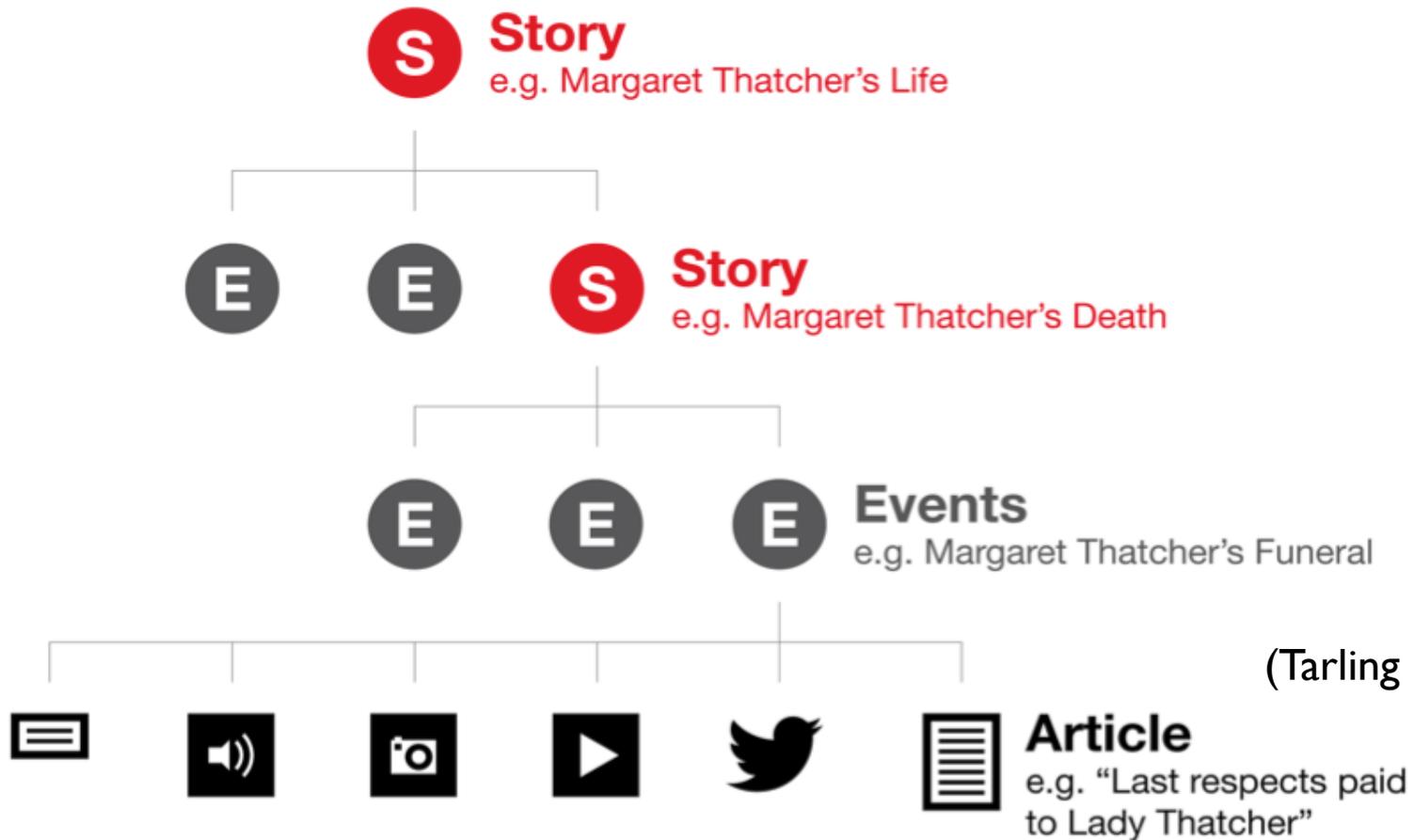
구조적 저널리즘(Structured Journalism)^{KGU}

▶ 구조적 저널리즘 (Structured Journalism)

- ▶ (Tarling 2013; 강정수 2014)
- ▶ 뉴스의 새로운 조직화 원칙
- ▶ 저널리즘 노동 과정에서 발생하는 정보의 가장 작은 단위를 분리하여 독립적인 객체(object)로 인식
- ▶ 이들 단위 객체를 묶어내고, 개별 단위를 지속해서 업데이트
- ▶ 새로운 객체 묶음을 생성하여 뉴스 스토리를 생성

구조적 저널리즘(Structured Journalism)

▶ 구조적 저널리즘 (뉴스 도달거리 확장 방안)

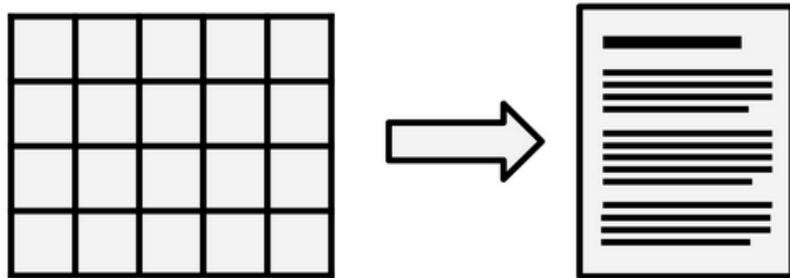


(Tarling 2013)

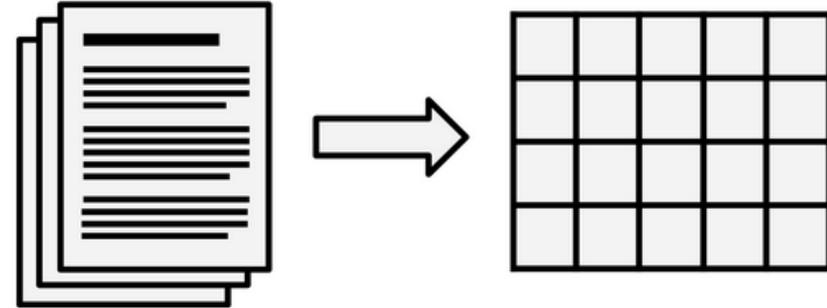
구조적 저널리즘(Structured Journalism)^{KGU}

▶ 구조적 저널리즘 - 데이터 저널리즘과의 차이

DATA-DRIVEN JOURNALISM



STRUCTURED JOURNALISM



<http://topdrawersausage.net/2013/07/07/storylines-vs-object-oriented-news/>

- ▶ 구조화된 저널리즘은, 개별 뉴스 정보(object)가 생성할 때 여기에 태그(tag)를 추가하여 (데이터베이스에) 저장하며, 이를 기초로 '새로운' 이야기를 만들어 내는 과정 (강정수 2014)

위기 탈출 전략 (Smart Contents)

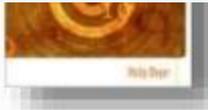
▶ Precision Journalism (Philip Meyer, 1969; 2002)



부족한 것은 기술적인 커뮤니케이션
수단이 아니라 **대중의 관심**이다.
Russell Neuman, <The Future of Mass Audience> 1991

그 관심은 **컨텐츠**에 달려있다.
필립마이어, <디지털 시대 저널리즘 구하기> 2004

Smart TV at CES: Delivering What
Consumers Don't Want?
January 09, 2013
TV manufacturers at CES have been touting their latest sets' interactive capabilities. But is that what consumers want?



14

▶ 저널리즘 정보 지식화 서비스

- ▶ 저널리즘 정보, 특히 뉴스기사 데이터베이스에 대한 통합 지식화 서비스는 아직 개척되고 있지 않은 생소한 분야임
- ▶ 지식화 서비스에 필요한 기반 기술들은 이미 많은 연구가 진행 중
 - ▶ 전문 용어 추출
 - ▶ 개체명 인식
 - ▶ 관계 추출
 - ▶ 문서 분류
 - ▶ 문서 요약
- ▶ 하지만 위의 기술들을 그대로 적용하기에는 아직 한계점이 있으며 잘 구성된 워크플로우를 바탕으로 한 **반자동 지식화 체제** 구축이 필요함

* 언론사의 정보를 재활용하는 각종 Buzz들은 제외

▶ 저널리즘 정보 지식화 서비스

▶ NewsMental (<http://www.newsmental.com/>)

▶ 기사 분석 서비스

- 핵심어 (인명, 지명, 기관명) 추출 및 기사 요약 서비스 제공

▶ 커뮤니티 오피니언 서비스

- 특정 기사에 대한 감성 및 영향력 분석을 사용자의 입력을 받아서 수동으로 처리함

▶ 분석 결과에 대한 패널 서비스 (대시보드 서비스) 기능 제공

▶ FAST (<http://fast.qcri.org/>)

▶ 페이지 뷰 횟수 예측 서비스

- 특정 기사에 대한 향후 페이지 뷰 (인기도) 예측 분석 수행
- 동적으로 기사 순위가 변동됨

▶ 저널리즘 정보 지식화 서비스

▶ NHN의 신문기사 지식화 서비스 추진 현황

▶ 뉴스 클러스터링 (<http://lab.naver.com/newscluster/>)

- 자동 클러스터링 기술을 이용해서 관련 기사를 통합
- 클러스터링 수행 시에 추출되는 키워드를 중요도에 따라 정렬
- Google에서는 이미 몇 년 전부터 서비스하고 있는 항목임

▶ 뉴스 라이브러리 (<http://newslibrary.naver.com/>)

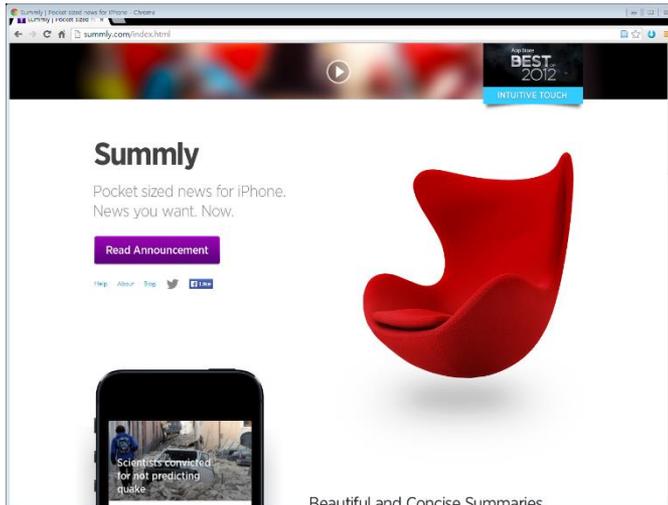
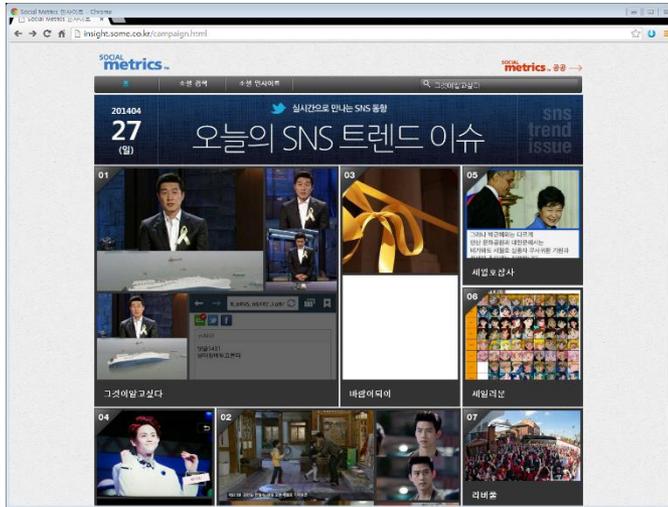
- 동아일보, 경향신문, 매일경제, 한겨레신문(최근 추가)이 창간 시점부터 1999년까지 발간한 신문 기사를 모두 디지털화 완료
- 통합 검색 서비스 제공 (날짜, 키워드, 상세 검색 지원)

국내외 뉴스미디어 지식 서비스 사례

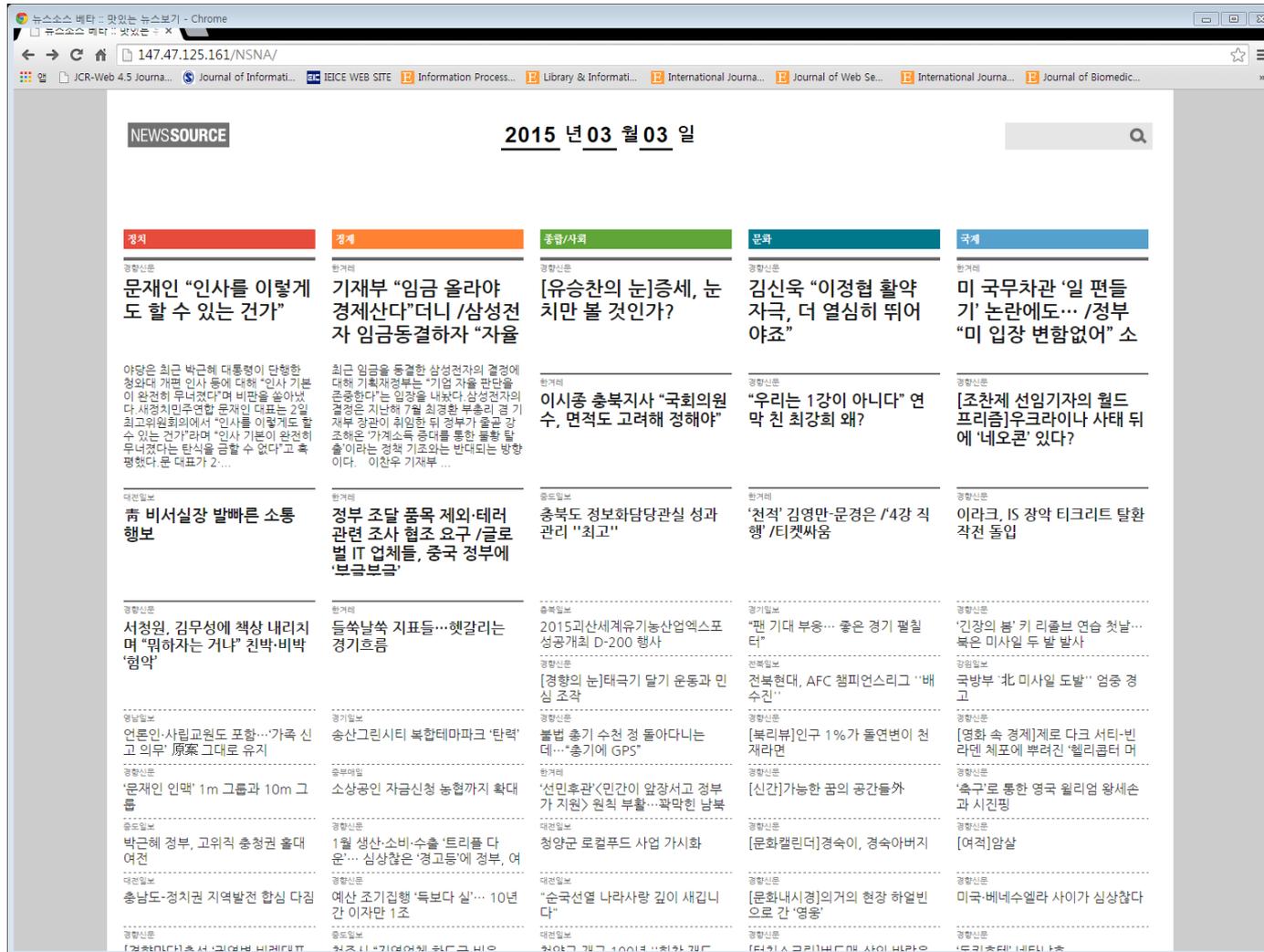
▶ 저널리즘 정보 지식화 서비스

- ▶ Social Metrics (<http://insight.some.co.kr/campaign.html>)
 - ▶ 다음소프트에서 운영하는 소셜 검색 및 분석 서비스
 - ▶ 특정 탐색어에 대한 소셜 미디어에서의 연관관계 서비스 제공
 - ▶ 탐색어에 대한 통계 추이 및 긍정/부정 추의 제공
 - ▶ 탐색어 여론 분석 정보 제공
- ▶ 위키트리 (<http://www.wikitree.co.kr>)
 - ▶ SNS를 이용한 뉴스 서비스 제공
- ▶ 씬리 (<http://summly.com/index.html>)
 - ▶ 자동요약을 이용해 뉴스를 3줄로 요약해서 모바일로 제공
 - ▶ 야후에 3,000만불에 M&A됨

국내외 뉴스미디어 지식 서비스 사례



국내외 뉴스미디어 지식 서비스 사례



위기 탈출 전략 (Smart Contents)

▶ Intelligent Journalism Process (신동희, 2014)

뉴스기사 지식베이스

① 데이터의 수집과 분석 (Digging deep into big data)



② 데이터의 정리와 핵심적 정보 추출 (Mining for nuggets of information)



③ 그래픽 비주얼화 (Visualizing information in graphics or multimedia specials)



④ 비주얼화한 데이터를 통해 스토리 추출 (Connecting classic storytelling with otherwise dry statistics)



⑤ 스토리를 통해 새로운 가치의 창출 (Creating media that has value for readers/users)

뉴스기사 분석서비스

▶ 정보 추출 (Information Extraction) 기술

- ▶ 특정 뉴스 기사의 내용을 이끌어가는 핵심적인 문구(linguistic expression)를 인식(recognition), 분류(classification), 식별(identification)하는 기술 (핵심 개체 추출)
- ▶ 식별까지 완료된 핵심 개체 간의 의미적 연관 관계를 추정
 - ▶ (핵심개체 - 관계 - 핵심개체)

검찰 수사관들이 28일 오후 유병언(73) 전 세모그룹 회장 차남 혁기씨 명의의 페이퍼컴퍼니인 '기술루션' 주소지인 서울 강남구 청담동의 한 건물에서 압수물품을 차량으로 옮기고 있다.



유병언 - 운영하다 - 세모그룹
 유병언 - 부자관계 - (유)혁기
 (유)혁기 - 운영하다 - 기술루션
 기술루션 - 위치하다 - 서울 강남구 청담동

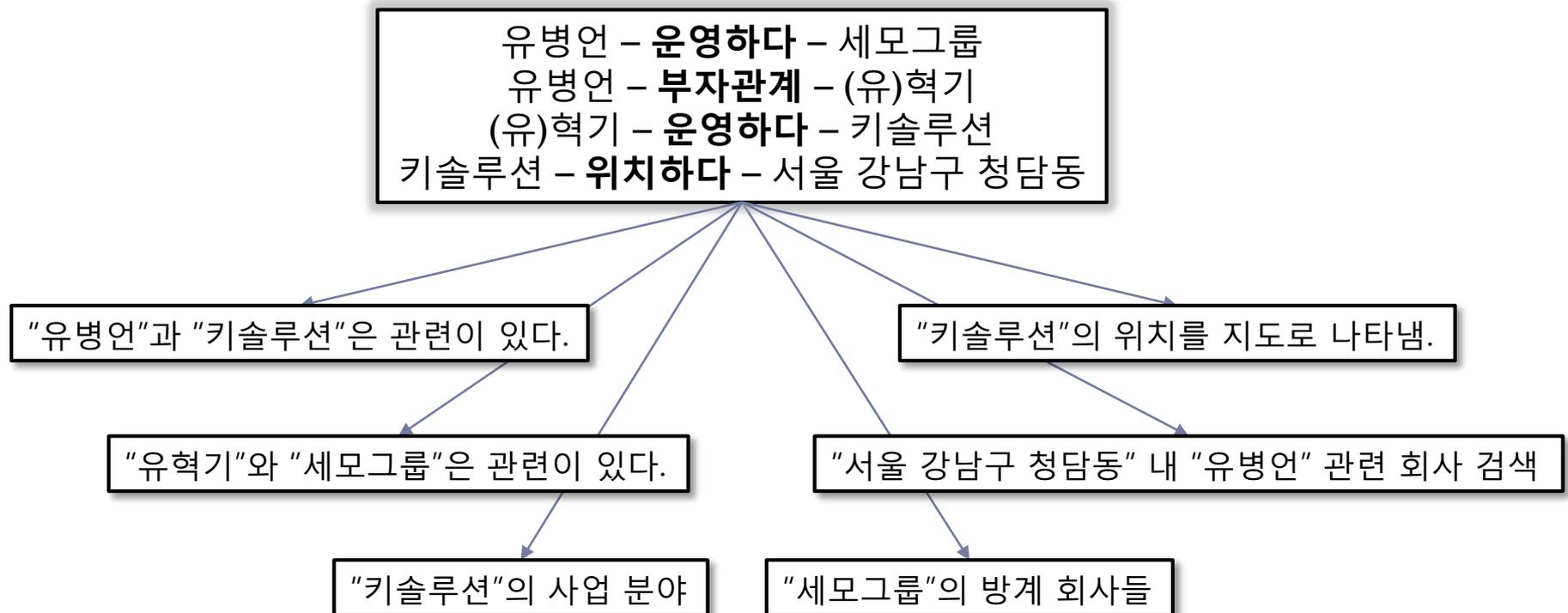
information
extraction

One stop solution to extract structured information from unstructured data.



▶ 정보 분석 (Information Analysis) 기술

- ▶ 정보 추출의 결과와 부가 정보를 바탕으로 다양한 형태의 심층적인 분석을 수행하고 가시화하는 기술

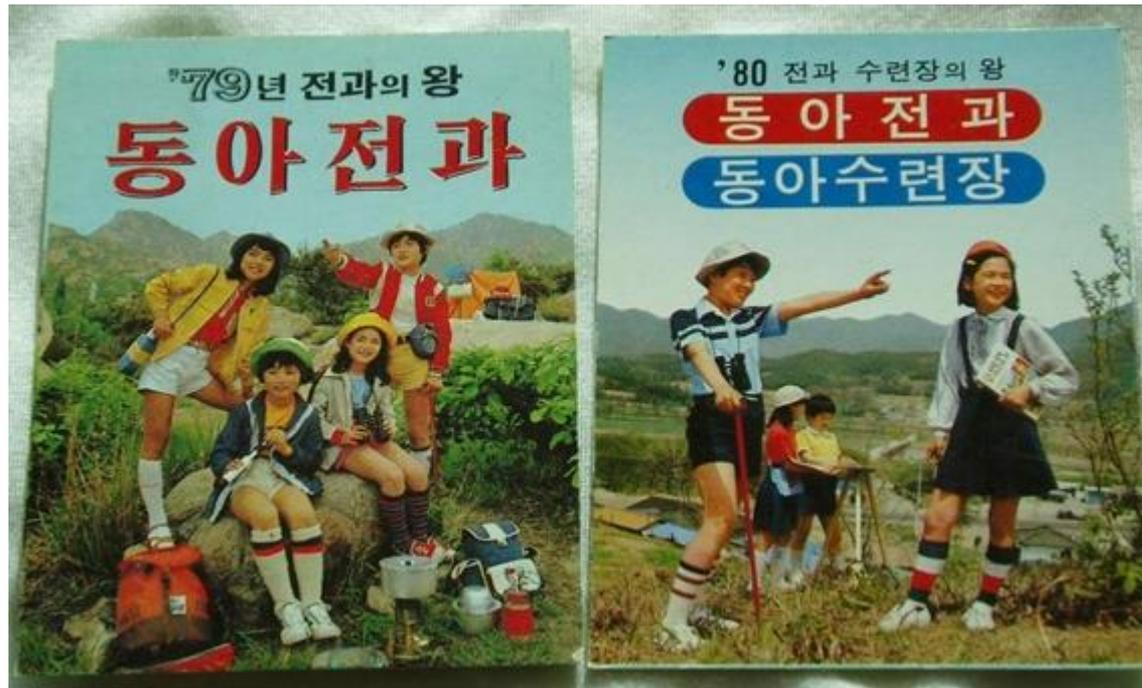


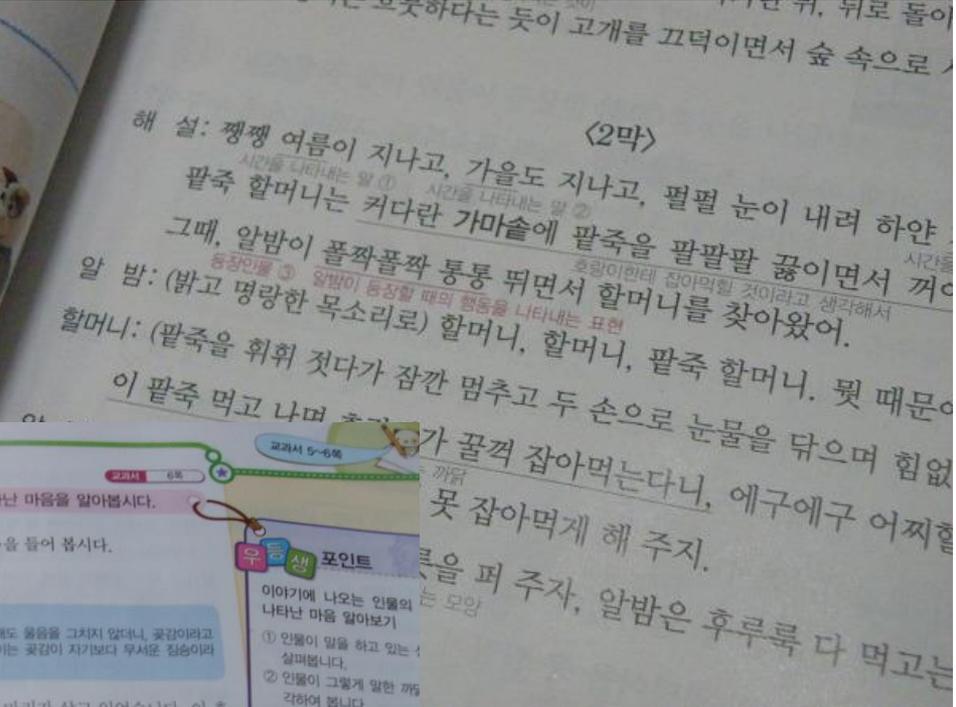
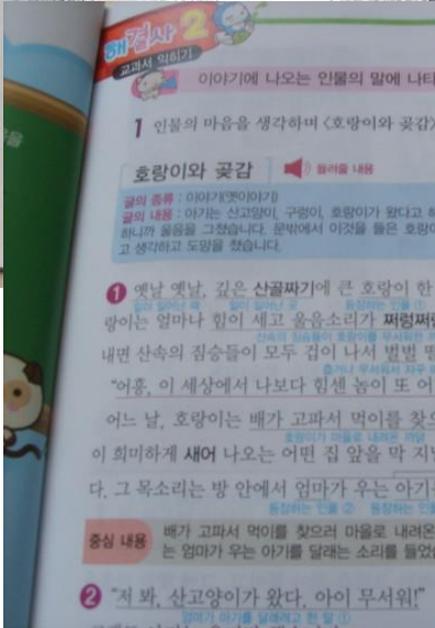
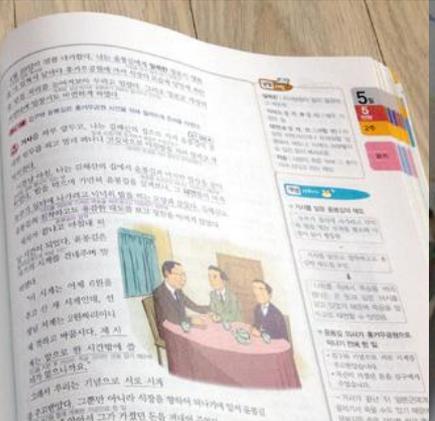
뉴스기사 지식베이스

▶ 도대체 어떻게 구축을 하지?



뉴스기사 지식베이스

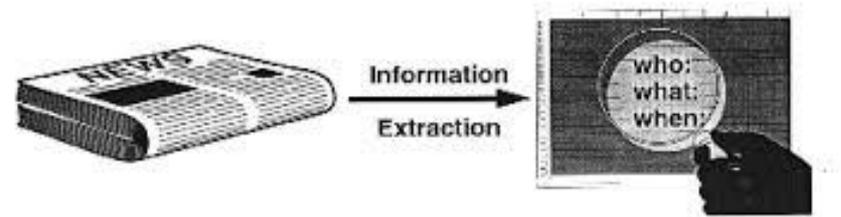




뉴스기사 지식베이스

▶ 정의

- ▶ “다년간”의 뉴스기사 데이터베이스를 “다각도”로 “추출”, “분류” 및 “요약”해서 정리한 고차원 데이터베이스
- ▶ 뉴스기사에 존재하는 중요한 문구들을 표지(Annotation)하고 분류(Classification)함으로써 구축됨
- ▶ 다년간 → 시간적 정보 (시계열 분석이 가능)
- ▶ 다각도 분류
 - ▶ 뉴스기사 내 행정구역 분류
 - ▶ 뉴스기사 내 기관/기업 분류
 - ▶ 뉴스기사 내 인물 분류
 - ▶ 장르 분류
 - ▶ 주제 분류
- ▶ 요약 추출
 - ▶ 키워드, 주제어, 핵심어, 요약문구



뉴스기사 지식베이스

1 중앙일보 뉴스 경제 | 라이프 | 오피니언 | 핫이슈 | 포토

검색

전체 정치 사회 지구촌 문화 지역 스포츠 연예 영미뉴스 블로그뉴스 뉴스클럽 보도자료 제보 일간스포츠

정치 | 최신기사 | 선거 | 국회·정당 | 청와대 | 외교 | 북한 | 국방 | 여론조사

리미엄 | 구독신청 | PDF

김한길-안철수, 신당 창당 전격 합의

[온라인 중앙일보] 입력 2014.03.02 10:34 / 수정 2014.03.02 11:42



김한길 안철수 신당 창당 [사진 = 중앙 포토 DB]

글자크기 +

시간: 2014.03

주제: 국회/정당

인물: 김한길, 안철수

기관: 민주당, 새정치연합, 신당

장르: 기사

면종: 1면 뉴스

행정구역: 서울 여의도

키워드: 신당 창당, 받겠다

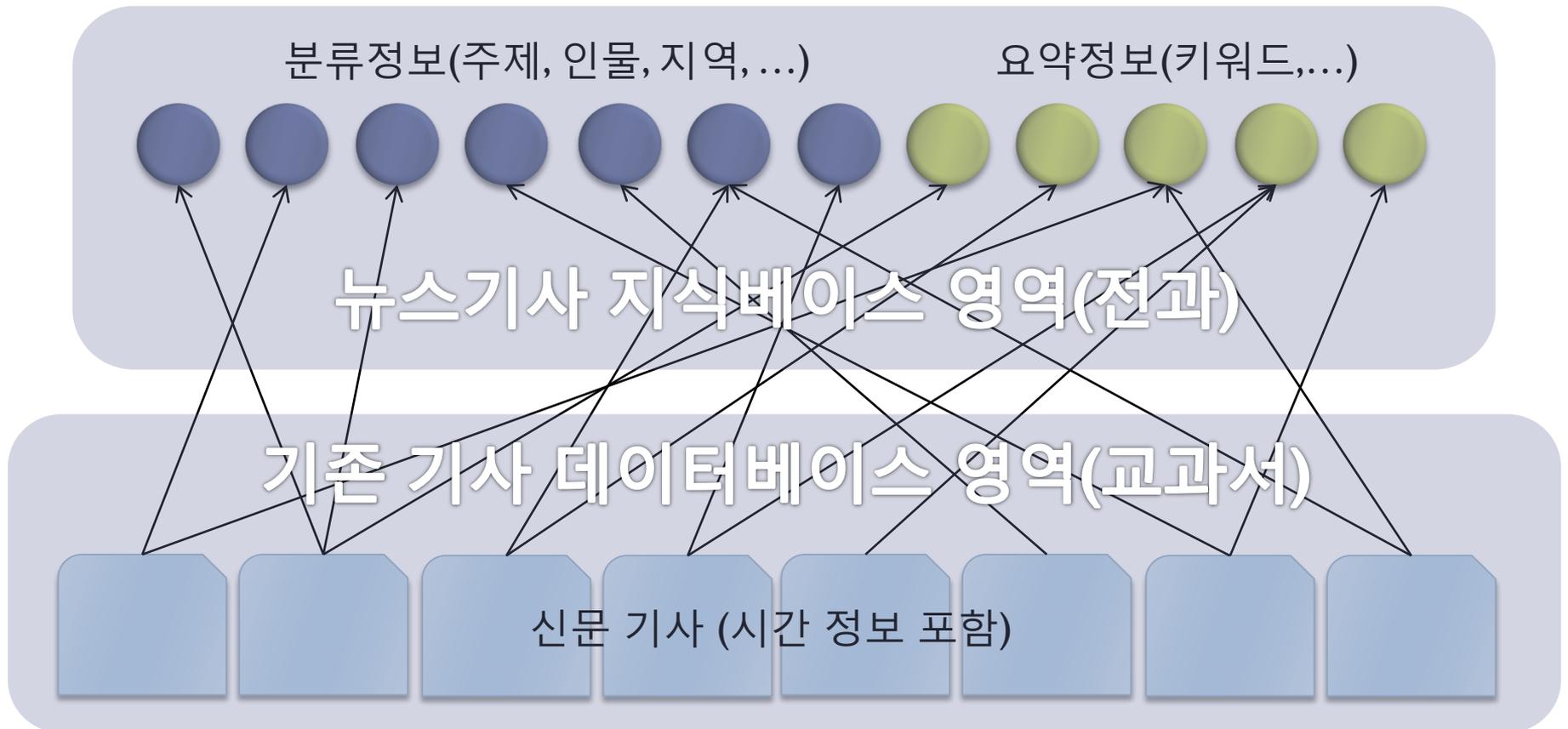
요약문구: 신당 창당 합의 미혼녀, 결

감성정보: 중립 고 평평 은 이유는

북 미사일 쏘짜마! 미 핵추진 잠수함 입항후

뉴스기사 지식베이스 – Mapping

단순 키워드가 아닌 토픽, 컨셉, 개념으로 접근
분석을 위해서 지식베이스 영역을 활용



뉴스기사 지식베이스 - 뉴스코드

뉴스표준

NewsML만의 새로운 소식과 다양한 정보



NewsML

NewsCode

- 한국형 뉴스코드(KsNewsCode)
- IPTC표준뉴스코드

한국형 뉴스코드(KsNewsCode)

KSOrgan(기관.단체)

KSPProduct(상품)

KSPeople(인물)

KSGenre(장르)

KSPageType(면종)

KSCCompany(기업)

KSRegion(행정구역)

KSSubjectCode(주제분류)

Character Set:

보기방식:



한국형 뉴스코드(KsNewsCode)의 추가 의견이나 건의사항이 있으면 검토 반영하도록 하겠습니다.

[▶ 건의/지원 요청하기](#)



토픽셋 : 기관및조직(ksOrgan)

N1TS :

- 현재 배포버전 : 2011-12-15
- 최초 배포버전 : 2009-05-22

G2KI :

- 현재 배포버전 : ,
- 최초 배포버전 :

설명 :

- 조직

다운로드 :N1TS,G2KI

뉴스기사 지식베이스 - 뉴스코드

▶ 뉴스코드를 이용한 지식베이스 구축

▶ 분류

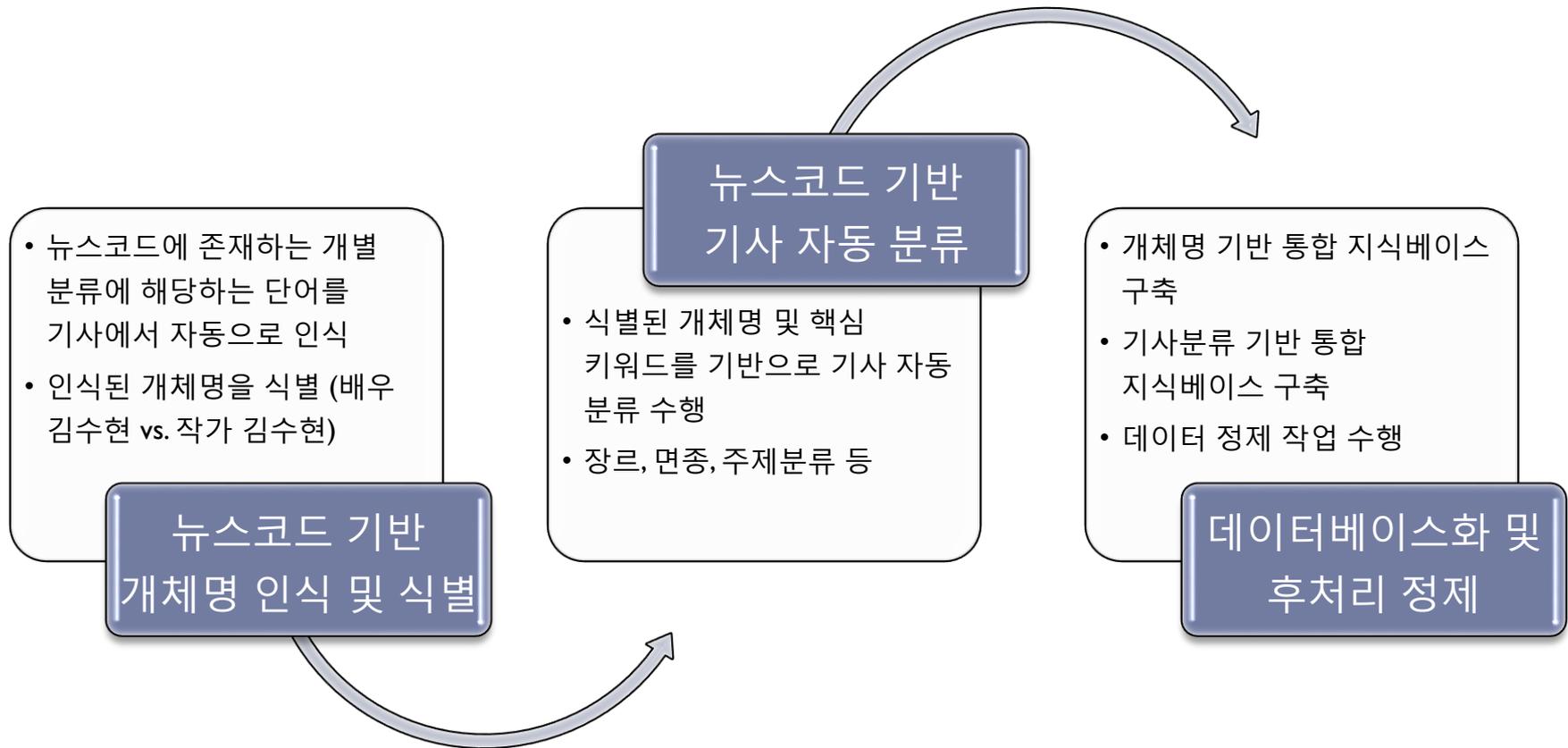
- ▶ 기관/단체
- ▶ 상품
- ▶ 인물
- ▶ 장르
- ▶ 면종
- ▶ 기업
- ▶ 행정구역
- ▶ 주제분류
- ▶ **감성분류**

▶ 요약

- ▶ 키워드, 핵심어, 중심어, 요약 문구, 인물에 대한 행위



뉴스기사 지식베이스 - 구축 방법



뉴스코드 기반 개체명 인식 및 식별

▶ 개체명 인식 (Named-Entity Recognition)

- ▶ 텍스트 내에 존재하는 인명, 지명, 기관명 등과 같은 고유명사를 자동으로 인식하여 분류하는 기술
- ▶ 국외 (영어) 수준
 - ▶ 상용화 수준에 도달했다고 많은 연구단체들이 주장하고 있으나 실제 체감 성능은 아직도 낮은 상황임. (특정 분야에 특화되어 최적화된 시스템들은 공개를 하지 않고 있음.)
 - ▶ 대표적 시스템들
 - Stanford NER, [UIUC NER](#), [LingPipe NER](#), ...
- ▶ 국내 (한글) 수준
 - ▶ ETRI 및 일부 기업 (다이퀘스트, 솔트룩스, 와이즈넷, 다음소프트 등) 중심으로 개발이 진행 중에 있음
 - ▶ 현재까지 만족할 만한 수준의 성능을 가진 시스템이 없음
 - ▶ 웹으로 테스트 가능한 시스템: [OWLNEST](#)

뉴스코드 기반 개체명 인식 및 식별

▶ 초기 시스템의 성능 개선 방안

- ▶ 개체명 인식 및 식별의 어려운 문제들
 - ▶ 기존 사전에 존재하지 않는 **신조어** 발굴
 - "세월호", "언딘인더스트리", "다이빙 벨", ...
 - ▶ 기존 사전에 존재하는 개체명에 대한 **동형이의어** 판별
 - "김수현", "유승민", "광주", "최재성", "신동아건설", "금강종합건설"
 - ▶ 기존 사전에 존재하는 개체명에 대한 **이형동의어** 판별
 - "삼성전자", "(주)삼성전자", "Samsung Electronics", "삼성", ...
- ▶ 학습 기반 시스템의 특징
 - ▶ 학습 집합에서 나타나는 개체명 자체 및 주변 자질을 활용하여 유사한 자질을 가진 부분을 개체명으로 인식함
 - "... 김수현 ... 드라마 ... 작가 ..."
 - "김수현"을 "인물"로 인식하고 "작가 김수현"으로 식별함
 - "... 경기도 ... 광주시 ..."
 - ▶ 사전에 존재하는 표제어에 대해서는 우선순위를 높여서 개체명으로 인식

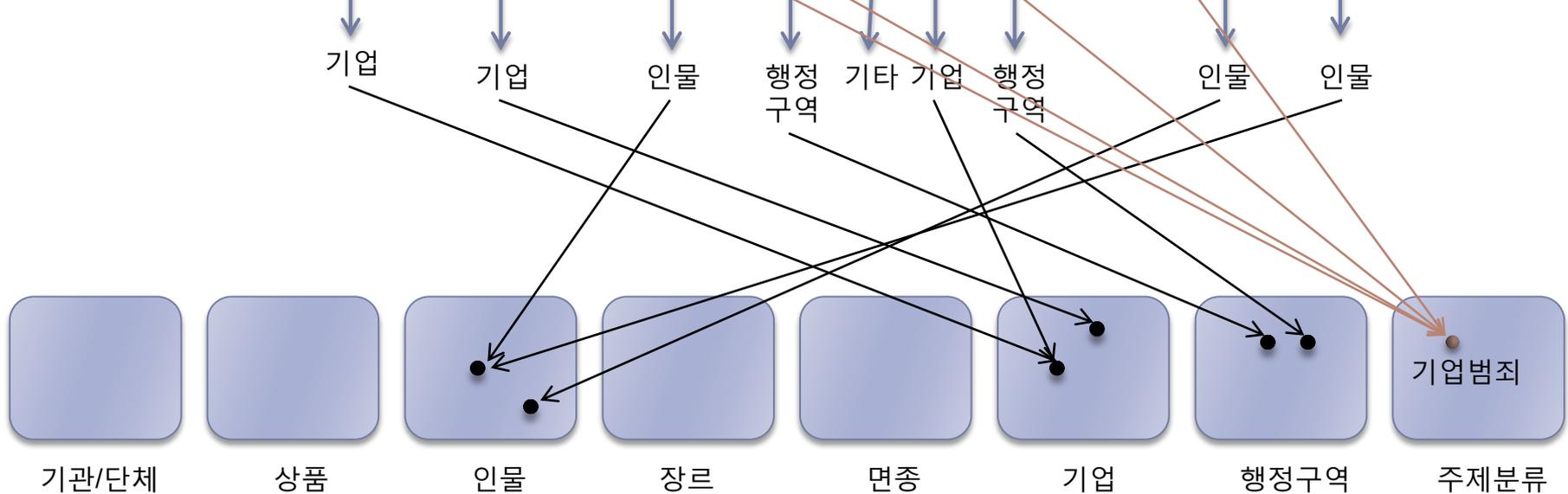
뉴스코드 기반 개체명 인식 및 식별

▶ 초기 시스템의 성능 개선 방안

- ▶ 가장 확실한 성능 개선 방법
 - ▶ 충분한 규모의 학습 집합 구축
 - ▶ 충분한 규모의 개체명 사전 구축
 - ▶ 학습 집합과 개체명 사전을 쉽고 편리하게 구축할 수 있는 도구 개발
 - ▶ 개발된 도구의 활발한 활용 및 지속적인 기능 강화를 통한 편의성 증대
 - ▶ 성능에 대한 자가 증강이 가능하도록 선순환적인 Boosting 기반 시스템 구성

뉴스코드 기반 개체명 인식 및 식별

검찰 수사관들이 28일 오후 **유병언**(73) 전 **세모그룹** 회장 차남 **혁기**씨 명의의 **페이퍼컴퍼니**인 '**키슬루션**' 주소지인 **서울 강남구 청담동**의 한 건물에서 압수 물품을 차량으로 옮기고 있다. 검찰은 이날 **세월호** 선사 실소유주인 **유병언** (73) 전 **세모그룹** 회장 일가의 **불법 외환거래**와 **비자금 조성** 혐의를 포착하고 관련 사무실 4곳에 대한 **압수수색**을 실시했다.



뉴스코드 기반 개체명 인식 및 식별



```

<?xml version="1.0" encoding="EUC-KR"?>
<document id="DEFAULT.d1" origId="DEFAULT.DID_00001">
  <sentence id="DEFAULT.d1.s0">
    <text><![CDATA[
세월호 선장과 선원 등 15명을 구속한 검경 합동수사본부가 수사의 초점을 세월호 선박 자체에 맞춰 사고원인에 대한 분석을 계속하고 있다.
]]></text>
    <entityList>
      <entity charOffset="1-3" id="
DEFAULT.d1.s0.e0" seqid="0" type="Ship_Name" name="세월호" nn="
세월호" from="manual" rule="" termhood="" term_hood_score="" audit="
true"/>
      <entity charOffset="23-31" id="
DEFAULT.d1.s0.e1" seqid="1" type="Institution" name="검경
합동수사본부" nn="검경 합동수사본부" from="manual" rule="" termhood=""
term_hood_score="" audit="true"/>
      <entity charOffset="42-44" id="
DEFAULT.d1.s0.e2" seqid="2" type="Ship_Name" name="세월호" nn="
세월호" from="manual" rule="" termhood="" term_hood_score="" audit="
true"/>
    </entityList>
  </sentence>
</chunkList>
</document>

```

7572 bytes 0.0 Processing

log chunk info ordered info predicate-argument info

[S1] ID: TESTBED [0] XML building ... C:\Users\spasie\google 드라이브\개발서술시스템 자료\과제\indi (성립교수님)\#indi\exam\news_new.xml
[S1] ID: TESTBED [0] XML end of job 1 C:\Users\spasie\google 드라이브\개발서술시스템 자료\과제\indi (성립교수님)\#indi\exam\news_new.xml

S0 ① 세월호 선장과 선원 등 15명을 구속한 검경 합동수사본부가 수사의 초점을 세월호 선박 자체에 맞춰 사고원인에 대한 분석을 계속하고 있다. ✓

S1 ② 검경 합수부는 선장 이준석 (69) 씨 등 선박직 15명이 전원 구속됨에 따라 앞으로는 사고의 구조적인 부분을 집중 수사하겠다고 28일 밝혔다. ✓

S2 ③ 수사총괄책임자인 안상돈 광주교검 차장검사는 이날 오전 광주지검 목포지청에서 기자들과 만나 이 같은 방침을 설명했다. ✓

S3 ④ 안 차장검사는 " (세월호의 화물) 과적 여부 , 고박 (고정) 적정성 , 증개톤 (증개축) 과정의 설계 및 시공 적정성에 대한 수사에 집중하겠다 " 고 말했다. ✓

S4 ⑤ 이어 " (세월호의) 안전설비 및 안전관리와 관련한 (업체) 여러군데를 압수수색하고 있다 " 면서 " 구명벌 등의 안전을 확인하기 위해 정비 및 검사 업체 관계자들도 소환해 조사 중 " 이라고 설명했다. ✓

S5 ⑥ 세월호는 15일 밤 인천에서 제주로 떠날 당시 복원력 (배가 한쪽으로 기울었을 때 원상회복하는 능력) 을 고려하지 않고 과적한 데다가 고박도 제대로 하지 않은 정황이 포착됐다. ✓

S6 ⑦ 여기에 무리한 증톤으로 무게중심이 위로 올라가 안전상에 문제가 있었다는 분석도 나오고 있다. ✓

S7 ⑧ 합수부는 초동대처가 부실했다는 질타를 받고 있는 해경과 관련해서도 목포해양경찰서 및 전남도소방본부 119상황실 , 진도 연안 해상교통관제센터 (VTS) 등을 압수수색하는 등 이 부분에 대한 수사도 하고 있다. ✓

S8 ⑨ 한편 합수부로부터 선장 이씨 등 3명의 신병을 인계받은 검찰은 허용범 전 중앙해양안전심판원 선임심판관을 전문수사 자문위원으로 위촉 , 수사에 참여토록 하기로 했다. ✓

[S1] ID: TESTBED [0] XML building ... C:\Users\spasie\google 드라이브\개발서술시스템 자료\과제\indi (성립교수님)\#indi\exam\news_new.xml
[S1] ID: TESTBED [0] XML end of job 1 C:\Users\spasie\google 드라이브\개발서술시스템 자료\과제\indi (성립교수님)\#indi\exam\news_new.xml

●●●●● sentence splitting ...
34.3 FPS

File : news_new.xml
Sentence : 11 items

뉴스코드 기반 개체명 인식 및 식별

Entity determination

File Entities

Enter text to search:

Reset highlight Search for selected Add as entity

REPORT DATE : 9 October 2008 [provided to CIA by Pakistani Criminal Investigation Unit , Karachi Division] NOTE : Surveillance report on the activities of Maulana Haq Bukhari , suspected to be a top leader within the Karachi faction of Lashkar-e-Jhangvi . Bukhari is frequently accompanied by Akram Basra , who acts as a driver and bodyguard . Additional information provided by police informants .

23 July 2008 - A delivery was made to a house in Lyari Town (a constituent town of Karachi) in a house believed to be used by Bukhari . The delivery was made by a two men in street clothing (as opposed to a uniform) who arrived in a white van , license LHR 6354 , with single blue stripe on each side . The delivery consisted of three medium boxes (requiring two hands to move) and a small box (handsized) . The boxes appeared to be heavy . One large box was square , the other rectangular . The small box was rectangular . It is unknown if Bukhari was home at the time of the delivery .

8 August 2008 - An unknown man visited the Lyari Town house where Bukhari is believed to stay . He arrived at 1615 , and was let into the house immediately . Loud voices could be heard for a few moments , then they subsided . About fifteen minutes later a silver Mercedes left the rear of the house with what appeared to be three occupants . Due to the tinted glass it was impossible to identify the occupants of the vehicle . The house was surveilled for the next five hours ; however no one came or went .

16 September 2008 - Bukhari was followed from his Lyari Town house to an apartment about 1.5 Kilometers southeast . He entered the building and stayed there for about two hours . When he left he returned to his Lyari Town house at which time the observer believed he 'd been discovered and left the area .

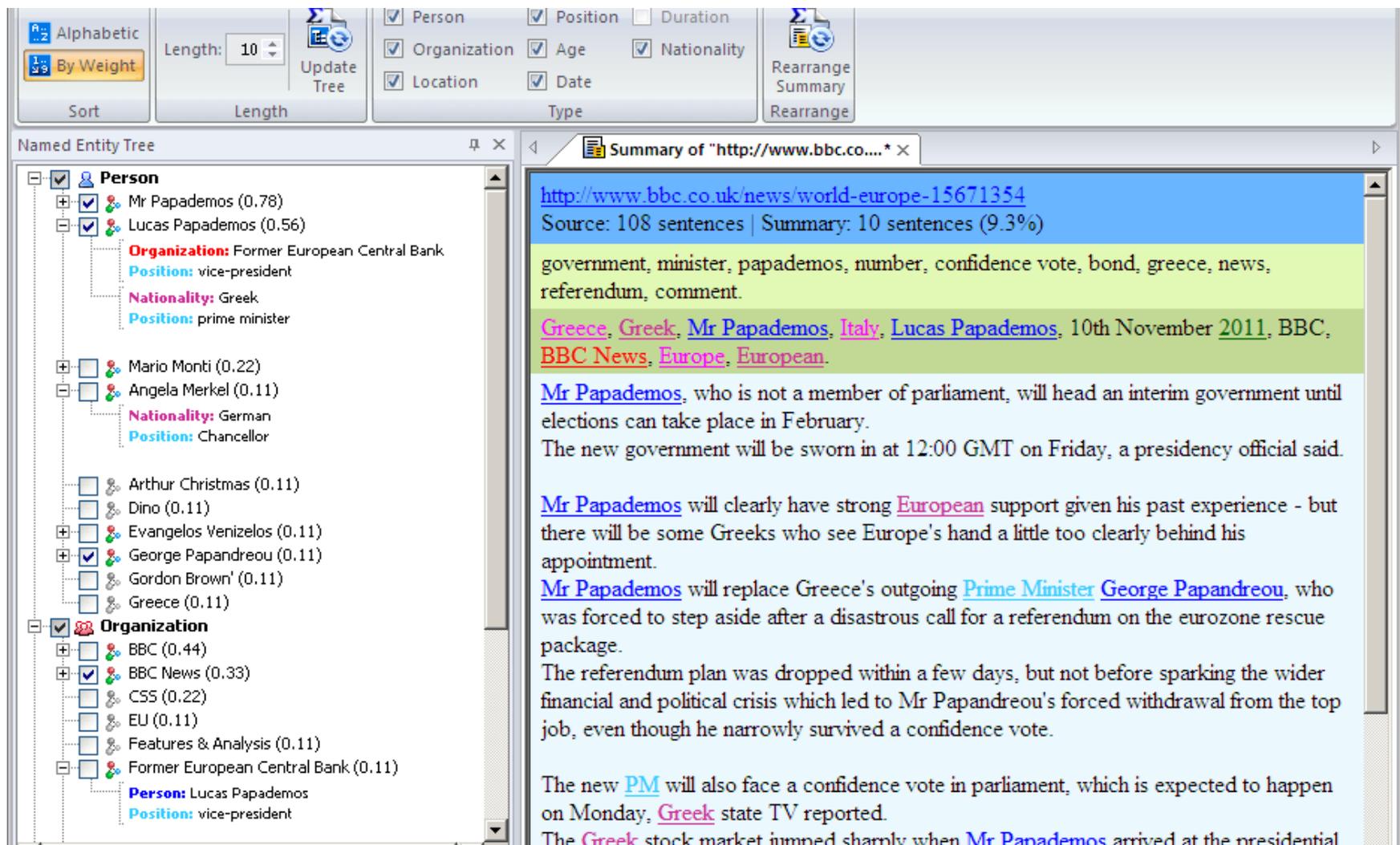
23 September 2008 - Bukhari and probably Basra are reported to have visited a house in the Katchi Abadis Old Settlement , on 835T Longhi Street . The informant , a vendor with business in the area , was passing by and noticed Bukhari and one other person enter the building at about 1430 . The informant knew Bukhari by sight , but had never met him .

0 entity found and highlighted.

Entity definition

Add entity			Remove entity			Move entity		
Contract entities			AutoContraction			Remove selections		
id	person_name	comment	id	organisation_n...	comment	id	MISC	comment
34	Tolya Shponar...	pilot;Belarus	16	Jhangvi		19	Ilyushin IL-76	cargo plane
36	Raph Plotnitsk...	pilot; Ukra...	14	CCTV		45	Karachi Division	
37	Pepik Malakho...	pilot;Ukrai...	32	Crime Suppres...				
38	Mykola Khitovo	pilot;Ukrai...	47	Maulana Haq ...				
39	Grigor Ialovskaia	pilot;Ukrai...	50	Bukhari				
42	Arkadi Borodi...	Ukraine;b...	55	Katchi Abadis ...				
51	Akram Basra							
id	location	city	country	id	tool	comment	id	date
29	Sri Lanka			4	Sixty National ...		43	February 11
35	Belarus			6	Computerized...		57	23 July 2008
40	Ukraine			61	LHR 6354	Mercedes	58	8 August 2008
46	Surveillance						59	16 September...
49	Lashkar-e-Jha...						60	23 September...
52	Lyari Town							
48	Karachi							
id	event	comment						

뉴스코드 기반 개체명 인식 및 식별



Sort: **Alphabetic** / By Weight | Length: 10 | Update Tree

Person: Person, Position, Duration
Organization: Organization, Age, Nationality
Location: Location, Date, Type

Named Entity Tree

- Person
 - Mr Papademos (0.78)
 - Lucas Papademos (0.56)
 - Organization: Former European Central Bank
 - Position: vice-president
 - Nationality: Greek
 - Position: prime minister
 - Mario Monti (0.22)
 - Angela Merkel (0.11)
 - Nationality: German
 - Position: Chancellor
 - Arthur Christmas (0.11)
 - Dino (0.11)
 - Evangelos Venizelos (0.11)
 - George Papandreou (0.11)
 - Gordon Brown' (0.11)
 - Greece (0.11)
- Organization
 - BBC (0.44)
 - BBC News (0.33)
 - CSS (0.22)
 - EU (0.11)
 - Features & Analysis (0.11)
 - Former European Central Bank (0.11)
 - Person: Lucas Papademos
 - Position: vice-president

Summary of "http://www.bbc.co.uk/news/world-europe-15671354"
Source: 108 sentences | Summary: 10 sentences (9.3%)

government, minister, papademos, number, confidence vote, bond, greece, news, referendum, comment.

Greece, Greek, Mr Papademos, Italy, Lucas Papademos, 10th November 2011, BBC, BBC News, Europe, European.

Mr Papademos, who is not a member of parliament, will head an interim government until elections can take place in February.
The new government will be sworn in at 12:00 GMT on Friday, a presidency official said.

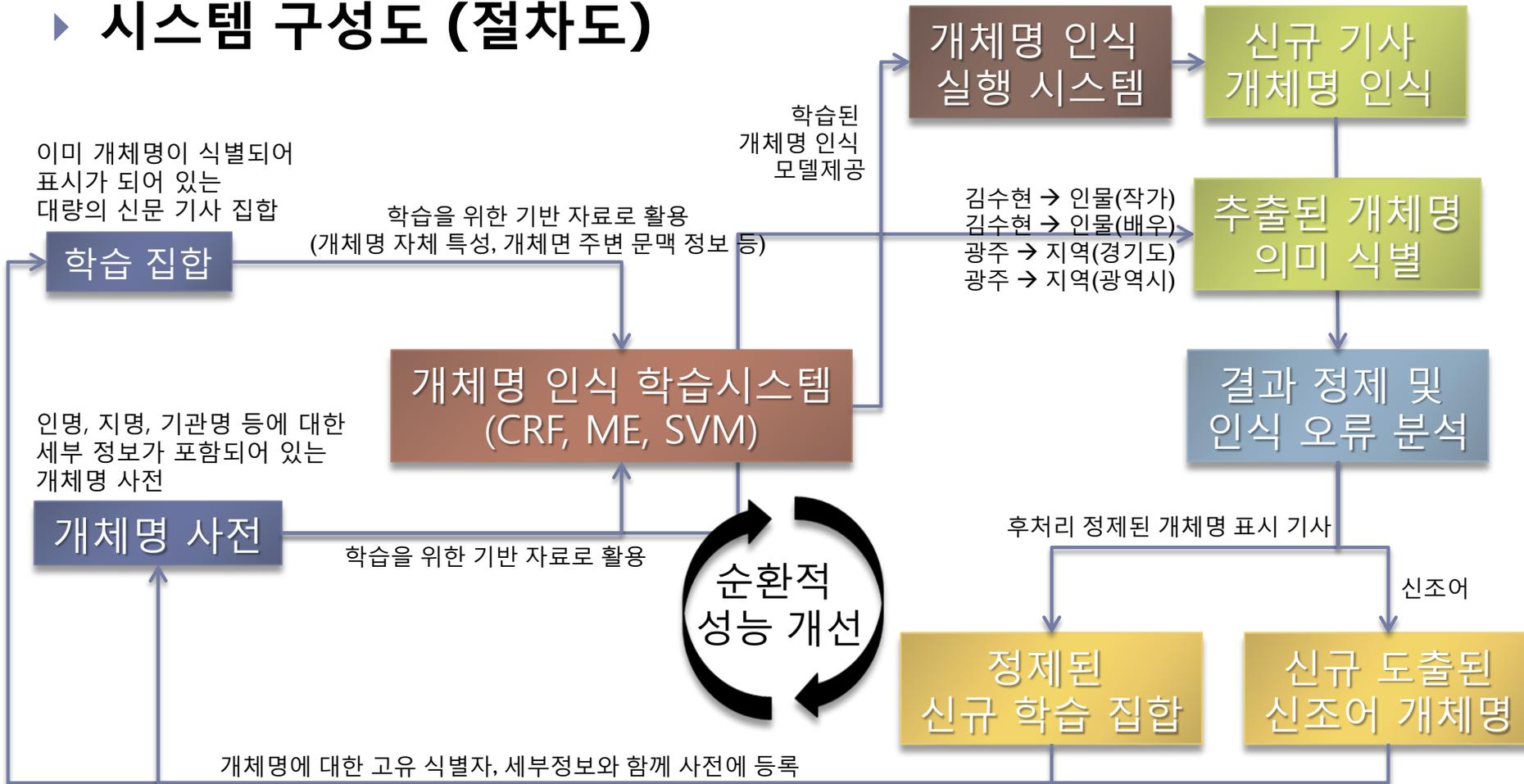
Mr Papademos will clearly have strong European support given his past experience - but there will be some Greeks who see Europe's hand a little too clearly behind his appointment.

Mr Papademos will replace Greece's outgoing Prime Minister George Papandreou, who was forced to step aside after a disastrous call for a referendum on the eurozone rescue package.
The referendum plan was dropped within a few days, but not before sparking the wider financial and political crisis which led to Mr Papandreou's forced withdrawal from the top job, even though he narrowly survived a confidence vote.

The new PM will also face a confidence vote in parliament, which is expected to happen on Monday, Greek state TV reported.
The Greek stock market jumped sharply when Mr Papademos arrived at the presidential

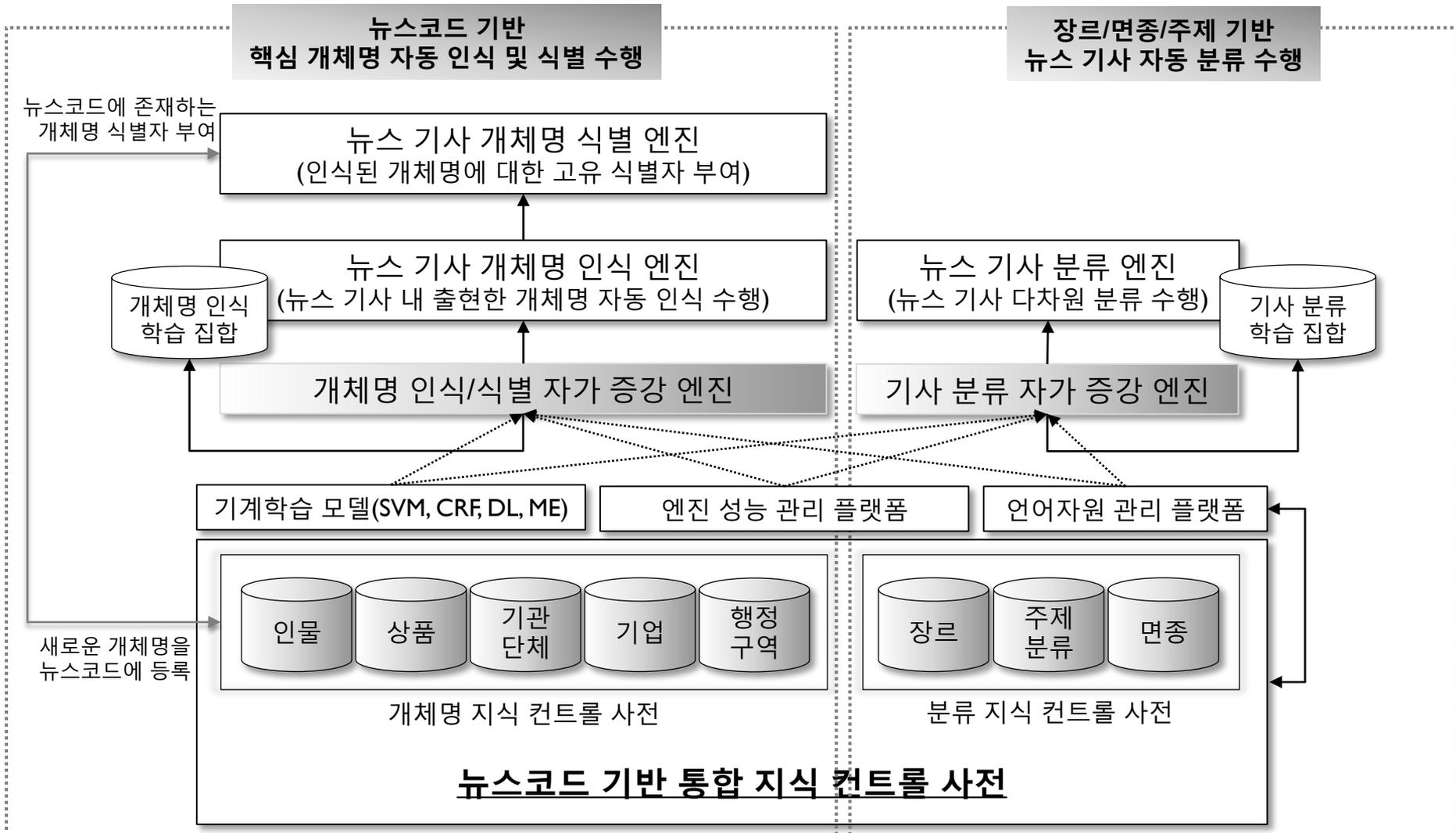
뉴스코드 기반 개체명 인식 및 식별

▶ 시스템 구성도 (절차도)



새롭게 분석된 신문 기사는 고품질의 학습 문서가 될 수 있으므로 이를 지속적으로 학습집합에 추가

뉴스코드 기반 지식베이스 시스템 (1/4) **KGU**



▶ 뉴스코드 기반 핵심 개체명 자동 인식 및 식별

▶ 개체명 자동 인식 엔진

- ▶ 뉴스 기사에서 개체명에 해당하는 문자열을 자동으로 인식하고 분류
 - (인물? 지명? 회사명?)

▶ 개체명 자동 식별 엔진

- ▶ 인식된 문자열 개체명을 뉴스코드 통합 사전을 이용하여 식별자 부여
 - (“김수현”(작가): 00012, “김수현”(배우): 00015)
- ▶ 뉴스코드 통합 사전에 존재하지 않으면 신규 등록

▶ 장르/면종/주제 기반 뉴스 기사 자동 분류

▶ 뉴스 기사 분류 엔진

- ▶ 뉴스코드에 등록된 장르별, 면종별, 주제분야별 자동 분류 수행
- ▶ 개별 뉴스기사에 분류코드를 부여

▶ 공통 모듈

- ▶ 기계학습 모델
 - ▶ 개체명 자동 인식/식별 및 기사 자동 분류에 사용되는 공통 기계학습 모듈 통합
- ▶ 엔진 성능 관리 플랫폼
 - ▶ 성능 관리자, 지식 구축자들의 편의성을 극대화한 사용자 인터페이스 탑재
 - ▶ 오류 검출, 신규 사전 엔트리 구축, 학습집합 강화 등의 기능 수행
- ▶ 언어 자원 관리 플랫폼
 - ▶ 뉴스코드 기반 통합 지식 컨트롤 사전 전체 관리
 - ▶ 엔진 성능 향상에 필요한 규칙베이스, 학습집합, 사전 등 관리

▶ 뉴스코드 기반 통합 지식 컨트롤 사전

- ▶ 기존에 구축된 뉴스코드 체제 유지
 - ▶ 개체명 코드
 - 인물, 상품, 기관/단체, 기업, 행정구역
 - ▶ 분류 코드
 - 장르, 주제분류, 면종
- ▶ 엔진을 통한 지속적 확장 수행
- ▶ 국각 언론정보 지식베이스 레지스트리 형태로 발전할 수 있는 다양한 기능 부여
 - ▶ 신규 엔트리 등록
 - ▶ 기존 엔트리 삭제/변경
 - ▶ Multi-tenant 기반의 뉴스코드 커스터마이징 기능 부여
 - 언론 기관마다 고유의 지식 컨트롤 사전 제공

추출 가능한 지식의 종류 (1/5)

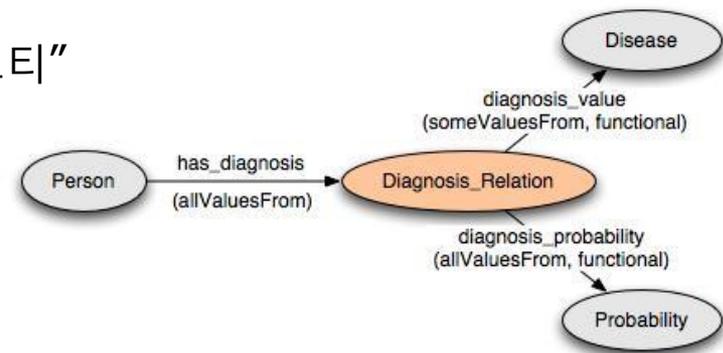
▶ 관계형 지식(용어 - 관계 - 용어)

▶ 세부 사항

- ▶ 전문용어, 인명, 지명, 기관명 인식이 가능해야 함.
- ▶ 전문용어, 인명, 지명, 기관명 간의 관계 식별이 가능해야 함.

▶ 예시

- ▶ “자동차 사고를 막을 수 있는 기술은 자율 응급 제동 시스템(AEB)이다.”
 - “자동차 사고” - “방지하다” - “자율 응급 제동 시스템”
 - “자율 응급 제동 시스템” - “동일하다” - “AEB”
- ▶ “불보자동차가 이런 기술을 세계 최초로 적용한 ‘시티 세이프티’ 기능을 실제로 선보였다.”
 - “불보자동차” - “출시하다” - “시티 세이프티”



추출 가능한 지식의 종류 (2/5)

▶ 절차적 지식 (A → B → C)

▶ 세부사항

- ▶ 해결해야 할 문제나 목표 문구 식별
- ▶ 문제 해결 및 목표 달성을 위한 절차 표현 식별

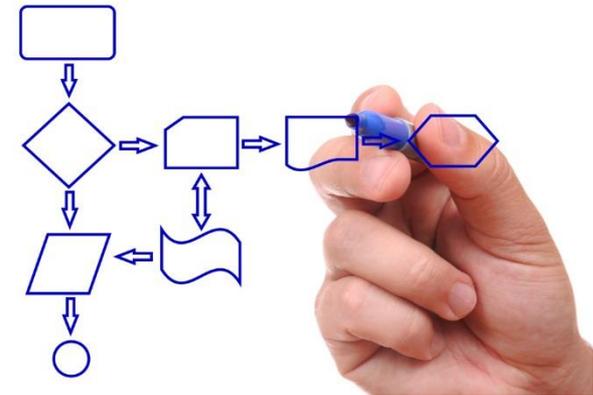
▶ 예시

▶ “과학캠핑 방법”

- (1)집터 고르기→(2)방위 정하기→(3)불 피우기→(4)주변 관찰하기

▶ “즉석 해시계 만드는 법”

- (1)평평한 땅에 짚막한 나무막대를 세우거나 꽃은 후 그림자의 끝부분에 표시
- (2)10분 후 다시 표시를 한 후 두 점을 연결해서 직선을 그려서 동-서 축 표시
- (3)동-서 축에 수직선을 그어서 남-북을 표시



추출 가능한 지식의 종류 (3/5)

▶ 정의형 지식 (“A”는 “B”로 정의할 수 있다)

▶ 세부 사항

- ▶ 특정 용어 및 개념에 대한 정의를 내린 문장을 식별

▶ 예시

- ▶ “**칼데라(caldera)**란 화산 폭발이 끝난 후 마그마가 빠져나와 생긴 지하의 빈 공간이 산 정상부의 무게를 이기지 못해 꺼져 생긴 넓은 함몰분화구를 말한다.”
- ▶ “법률적으로는 인공조명의 부적절한 사용으로 인해 과도한 빛이 생기거나 정해진 영역 밖으로 누출되는 빛이 건강하고 쾌적한 생활을 방해하는 상태를 **빛공해**로 규정한다.”
- ▶ “**횡문근**은 팔이나 다리 등 움직이는 부위에 붙어 있는 가로무늬 근육이다.”
- ▶ “**위상수학**을 간단히 말하자면 공간 속의 점, 선, 면 그리고 위치 등에 관하여 양이나 크기와는 상관없이 형상이나 위치 관계를 나타내는 수학의 한 분야이다.”

추출 가능한 지식의 종류 (4/5)

▶ 수치로 표현된 통계적 지식

▶ 세부 사항

- ▶ 문헌에 표현된 각종 지표, 통계, 수치 등을 추출하여 정형화

▶ 예시

- ▶ “가장 앞선 기술력을 자랑하는 미국이 운용하는 군용 무인기 비율은 **2005년 5%에서 2010년 41%로 크게 늘어났다.**”
- ▶ “항공 우주 관련 연구 기관인 팀 그룹(Teal Group Corp.)은 **2023년 전 세계 무인기 시장 규모가 890억 달러(약 93조 원) 수준으로 성장할 것**이라는 전망을 내놨다.”
- ▶ “고고도(高高度, **지상 7,000에서 10,000m쯤의 높이**)”
- ▶ “영국 방위 산업체 키네틱 사의 무인기 ‘제퍼(Zephyr)’가 지난 **2010년 7월 14일 21분 연속 비행**한 기록을 가지고 있다.”

추출 가능한 지식의 종류 (5/5)

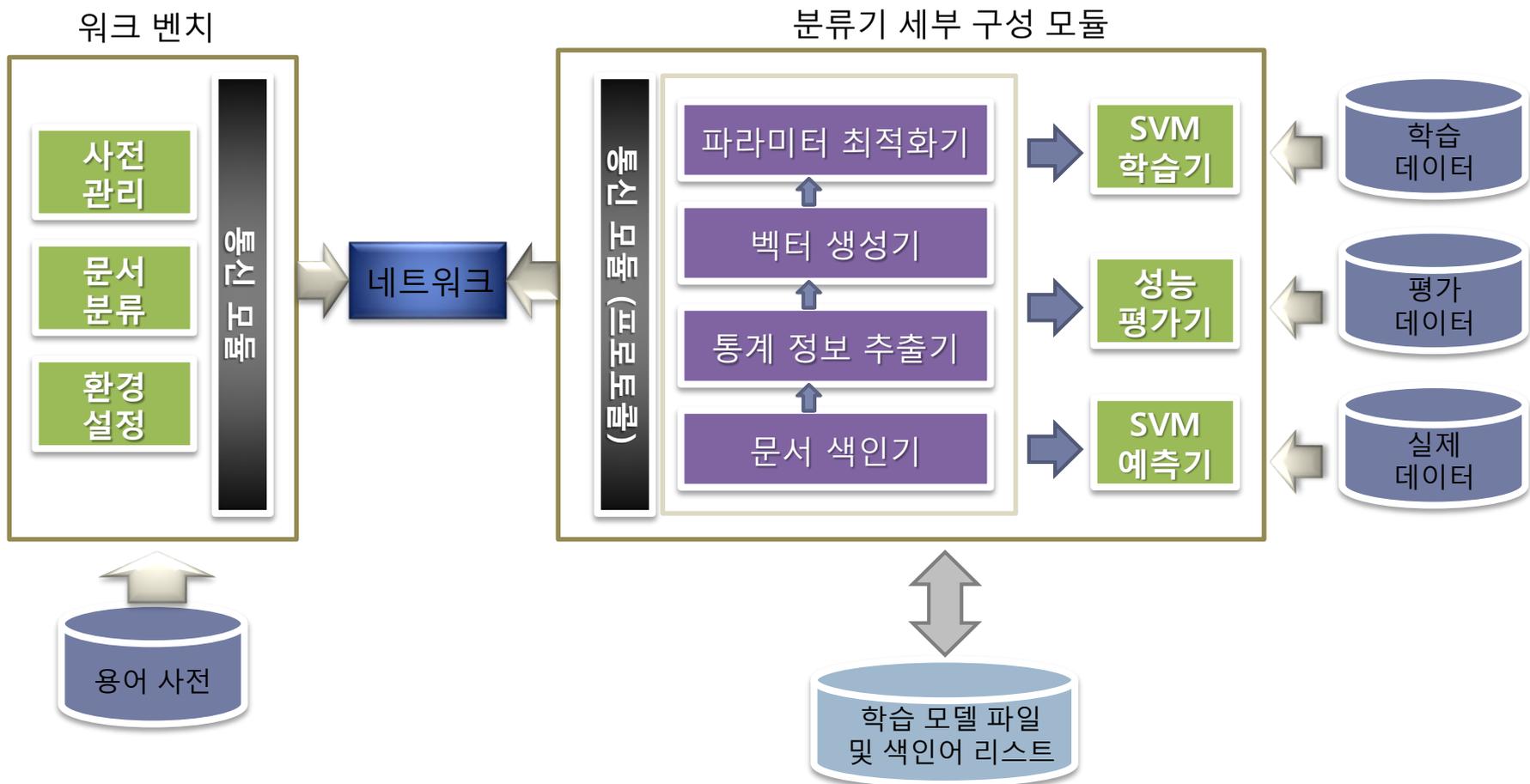
▶ 기타

- ▶ 그 외에도 다양한 형태의 지식을 추출할 수 있음
 - ▶ 인물의 신상명세
 - ▶ 종류 (작은빨간집모기, 중국얼룩날개모기, 아에데스 알보픽투스)
 - ▶ 사건의 진행 과정 (임진왜란 진행 과정 등)

- ▶ 텍스트를 수동으로 분석하면서 다양한 형태의 지식을 확인하고 이를 정형화하는 방안이 필요함
- ▶ 동시에 콘텐츠 이용자에게 어필할 수 있는 서비스를 함께 고민함으로써 서비스 중심의 지식베이스 구축을 통한 활용성 극대화를 고려해야 함

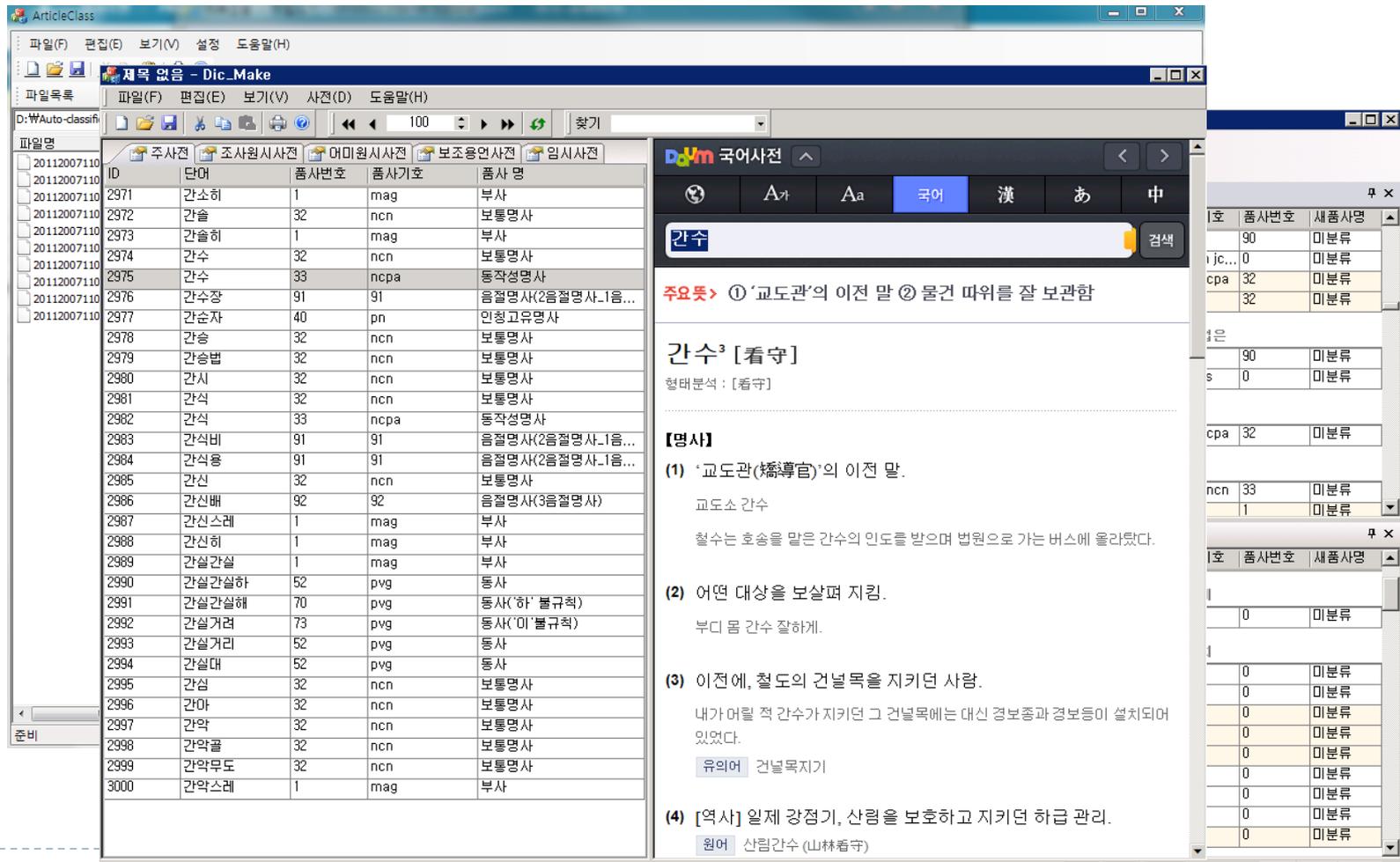
뉴스코드 기반 기사 자동 분류

▶ 기사 자동 분류 시스템 구성도



뉴스코드 기반 기사 자동 분류

▶ 한국어 텍스트 분석 지원 도구



The screenshot displays a software interface for text analysis. On the left, a window titled 'Dic_Make' shows a list of words with their corresponding codes and names. The main window shows a search for '간수' in the 'Dynam' Korean dictionary, providing detailed information about the word's usage and classification.

ID	단어	품사번호	품사기호	품사명
2971	간소히	1	mag	부사
2972	간술	32	ncn	보통명사
2973	간술히	1	mag	부사
2974	간수	32	ncn	보통명사
2975	간수	33	ncpa	동작성명사
2976	간수장	91	91	음절명사(2음절명사_1음...
2977	간순자	40	pn	인칭고유명사
2978	간승	32	ncn	보통명사
2979	간승법	32	ncn	보통명사
2980	간시	32	ncn	보통명사
2981	간식	32	ncn	보통명사
2982	간식	33	ncpa	동작성명사
2983	간식비	91	91	음절명사(2음절명사_1음...
2984	간식용	91	91	음절명사(2음절명사_1음...
2985	간신	32	ncn	보통명사
2986	간신배	92	92	음절명사(3음절명사)
2987	간신스레	1	mag	부사
2988	간신히	1	mag	부사
2989	간실간실	1	mag	부사
2990	간실간실하	52	pvg	동사
2991	간실간실해	70	pvg	동사('하' 불규칙)
2992	간실거리	73	pvg	동사('미' 불규칙)
2993	간실거리	52	pvg	동사
2994	간실대	52	pvg	동사
2995	간심	32	ncn	보통명사
2996	간마	32	ncn	보통명사
2997	간약	32	ncn	보통명사
2998	간약꿀	32	ncn	보통명사
2999	간약무도	32	ncn	보통명사
3000	간약스레	1	mag	부사

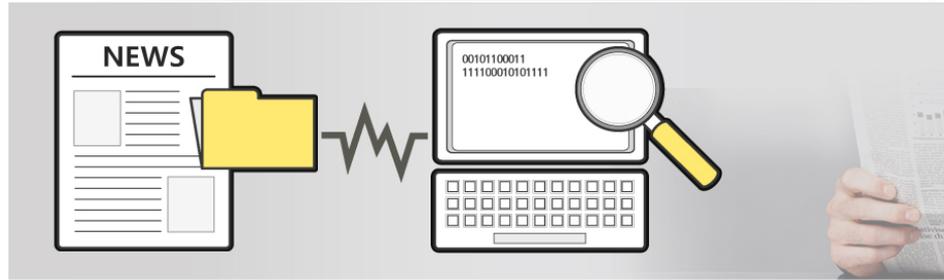
간수 [看守]

형태분석: [看守]

【명사】

- '교도관(矯導官)'의 이전 말.
교도소 간수
철수는 호송을 맡은 간수의 인도를 받으며 법원으로 가는 버스에 올라탔다.
- 어떤 대상을 보살피 지킴.
부디 몸 간수 잘하게.
- 이전에, 철도의 건물목을 지키던 사람.
내가 머물 적 간수가 지키던 그 건물목에는 대신 경비종과 경비등이 설치되어 있었다.
유의어: 건물목지기
- [역사] 일제 강점기, 산림을 보호하고 지키던 하급 관리.
원어: 산림간수(山林看守)

신문기사 자동분류 시스템



약 2만4천권의 기사 데이터로 학습된 문서 분류 시스템이며 입력된 기사를 분석하여 IPTC 360여개의 중분류중 신뢰도가 높은 5개의 결과값을 표시합니다.



서울시스템이 KISTI의 기술지원을 받아 개발중인 자동분류 시스템입니다.

신문기사 입력 (전체)

미국 최대 소비자단체인 소비자연합 (Consumers Union-CU)이 최근 미국에서 발생한 광우병이 예전에 발생했던 광우병보다 훨씬 더 위험하다고 경고하고, 미국과 한국 정부에도 적절한 대책을 세우라고 충고했다.

소비자시민모임 (소시모)은 "미 소비자연합 (CU)이 최근 전자우편을 통해 지난달 25일 미국에서 발생한 광우병 사건은 이전의 어떤 경우보다도 더 심각하며, 이는 과거 영국에서 발생했던 광우병 (classical BSE)과 달리 비전형적인 광우병 (atypical BSE) 계통이라고 알려졌다"고 2일 밝혔다.

소시모의 설명을 들어보면, 미 소비자연합의 수석연구원이자 '환경을 걱정하는 과학자들의 모임 (UCC)'에 소속된 광우병 전문가 마이클 한센 박사는 "미 정부가 공식적으로 발표하지는 않았지만, 국립수의연구소 (NIVSL)는 이번 광우병을 1-타입 계통으로 확인했다"며 "1-타입은 기존 광우병보다 더 빨리 감염되고 더 치명적 (more virulent)일 수 있다"고 경고했다.

미 소비자연합은 지난 1일 "미국의 광우병 대책이 적절치 못하다"며 미 농무부 (USDA)와 식품의약국 (FDA)에 "새로운 1-타입 광우병에 대해 완전하고 충분한 조사를 하고, 광우병 관리 감독 및 대책을 강화하라"고 요구했다. 이어 한국 정부에도 "이러한 내용을 참고해 미국 쇠고기 수입 정책에 반영할 것"을 조언했다.

소시모는 "정부는 미국산 쇠고기 수입을 즉각 중단하고 이번 광우병 사건의 전말을 밝히라"며 "향후 미국으로부터 쇠고기를 수입하게 될 경우에는 미국내에서 광우병 테스트를 받았다는 표시가 있는 쇠고기만을 수입하라"고 촉구했고, 이런 내용을 담은 서한을 1일 농림수산식품부에 제출했다.

자동분류시작

지우기

보건위생 : 질환 (분류코드:07001, 21%)

경제/산업 : 거시경제 (분류코드:04008, 13%)

경제/산업 : 농림수산업 (분류코드:04001, 4%)

경제/산업 : 기업정보 (분류코드:04016, 3%)

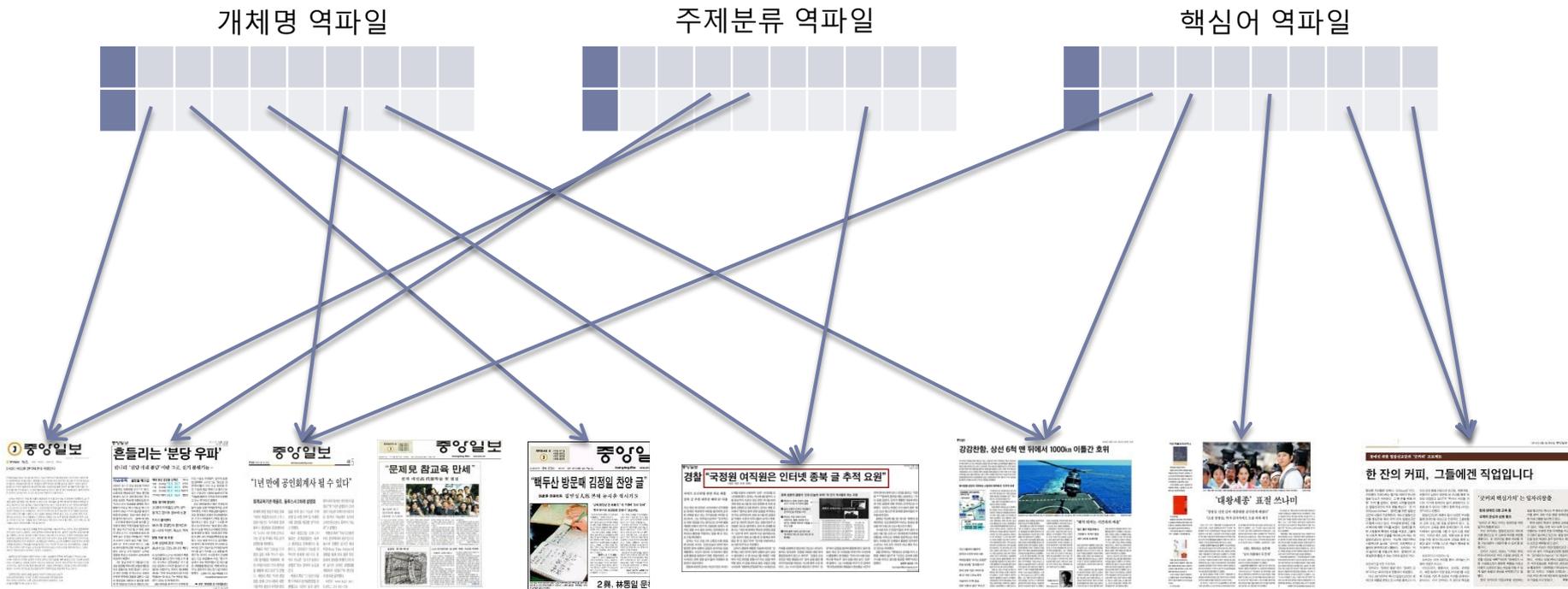
경제/산업 : 생활용품 (분류코드:04007, 3%)

데이터베이스화 및 후처리 정제

▶ 데이터베이스 설계

- ▶ 원본 신문기사에서 도출되는 부가 정보
 - ▶ 개체명 리스트, 주제분류, 핵심어 (+ 색인어)

개체명, 주제분류, 핵심어를 결합한 고수준의 질의가 가능함



뉴스기사 지식베이스의 효과

▶ 미시적 효과

- ▶ 지능형 질의 (Intelligent and Qualitative Queries)
 - ▶ “김한길과 안철수가 통합 합의를 하고 있을 때 **윤여준**은 어디서 무엇을 하고 있었나?”
 - ▶ “통합 합의 후 **청와대**의 반응은?”
 - ▶ “통합 합의 후 **미국**의 반응은?”
 - ▶ “통합 합의 이전의 **안철수**와 **김한길**의 행보는?”
 - ▶ “통합 합의 이전의 **새누리당**의 행보는?”
- ▶ 사건 사이의 숨겨진 관계 파악 (Tacit Relations)
 - ▶ “러시아의 크림반도 진격” : “우크라이나의 분열 사태”
 - ▶ “아베 총리 망언 및 망발” : “일본 내부 상황”
- ▶ 여론 분석 (Public Opinion Analysis)
 - ▶ “신당 창당 합의에 대한 여론 (좋다/싫다)”

뉴스기사 지식베이스의 효과

▶ 미시적 효과

- ▶ 통합 분석 및 융합 정보 제공
 - ▶ 지난 50년 간의 국내 선박 사고 일지
 - ▶ 미국 대통령 방한 전후의 국내 대형 사건/사고
 - ▶ 서울시 행정구역 중에서 기업 범죄가 가장 많이 발생하는 곳
 - ▶ 현재까지의 기업 범죄 현황

	기업 명칭	소유주	기업범죄방법	기소여부	처벌수위
...
...					
...					
...					

- ▶ 지난 20년 간 국내 선박 사고에 대한 정부 대응 방법
- ▶ 유병언 회장의 일대기 (활동)
- ▶ 김수현(연기자)의 (활동)

뉴스기사 지식베이스의 효과

Joins 검색

김수현



통합검색

뉴스

화보

이미지

메거진

PDF

인물

블로그

인물정보(총 28건)

더보기 >



김수현 상세정보보기 >

이름 김수현
생년 1943년
학력 -
소속 -

상세 정보량 기본0건, 학력0건, 경력0건, 기타0건

동명이인



김수현
1962년 | 공무원



김수현
1963년 | 금융인



김수현
1963년 | 언론인



김수현
1970년 | 반조인

많이 본 기사

- 1 손석희, 진도항서 나올때...
- 2 [포드앤]손태영 '검은 의
- 3 배 안에 몰 안 찼는데...
- 4 요즘 뜨는 제주도 독채 펜
- 5 세월호 이름 짓고 출항할
- 6 이영애, 김정 한복에 노란
- 7 5월엔 참을 수 없는 여
- 8 박진영 '아무 애가나 막
- 9 에디터 4인이 알려주는 봄
- 10 건강과 맛, 모두 만족하는

인물 기반의 정확한 시맨틱 검색도 가능



[\[세월호 침몰\] 김수현 이어 전지현까지 기부...후후후 '별'대 카탈](#) [아주경제] | 2014년04월29일(화) 오후 3:09

별'대 전지현 김수현 [사진제공=SBS] 아주경제 안전영 기자 = 지구인과 외계인의 사랑 이야기로 큰 인기를 얻은 '별에서 온 그대, 김수현 전지현 커플이, 통 큰; 기부로 이목을 집중시켰다. 선...

[아주경제]내재검색 | [관련뉴스검색]



[하지원, 기황후' 누물의 총영 기승남을 떠나보낼때](#) [이데일리SPN] | 2014년04월29일(화) 오후 3:06

→ 김수현의 그녀 안젤리베이라, 사랑스러운 패션 화보 → 소녀시절 "사실 그룹명은 어인천하, 여성시대, 부녀회장.." 폭소 <저작권자©함께 만들고, 함께 즐기는 엔터테인먼트 포털 이데일리 스타in> ...

[이데일리SPN]내재검색 | [관련뉴스검색]

뉴스기사 분석서비스 - 현재



The screenshot displays the Kinds.or.kr website interface. At the top, there's a search bar with '통합검색' and '김영란법' entered. Below the search bar, there are navigation tabs for '기사통합검색 Kinds', '언론전문자료', '언론정보광장', and '미디어링크'. The main content area shows search results for '김영란법' with a total of 5,202 articles and 0 high-quality articles. The results are categorized into '기사통합검색(5,162건)', '언론전문자료(38건)', and '언론정보광장(2건)'. Each category lists specific news items with titles, dates, and brief descriptions. For example, under '기사통합검색', there are articles from 2015.03.04 and 2015.03.04. Under '언론전문자료', there are articles from 2015.03.02, 2015.02.28, and 2015.02.26. Under '언론정보광장', there is an article from 2004.07.29.

뉴스기사 분석서비스 - 구성도

▶ News Article Analysis Service



서비스 제공 (HTML5)

기사 지식베이스 브라우저
(동적 분석 View 생성엔진)

가시화 방법에 따른
지식표현 방법 확장

지식 제공

외부 확산

지식 저장



외부에 제공



기사 콘텐츠
분석 및 지식 추출 (수동)

신규 지식 발견 및 스키마 확장

기사 콘텐츠
분석 및 지식 추출 (엔진)

학습집합 및 기반 사전으로 활용

뉴스기사 분석서비스 - 모듈화

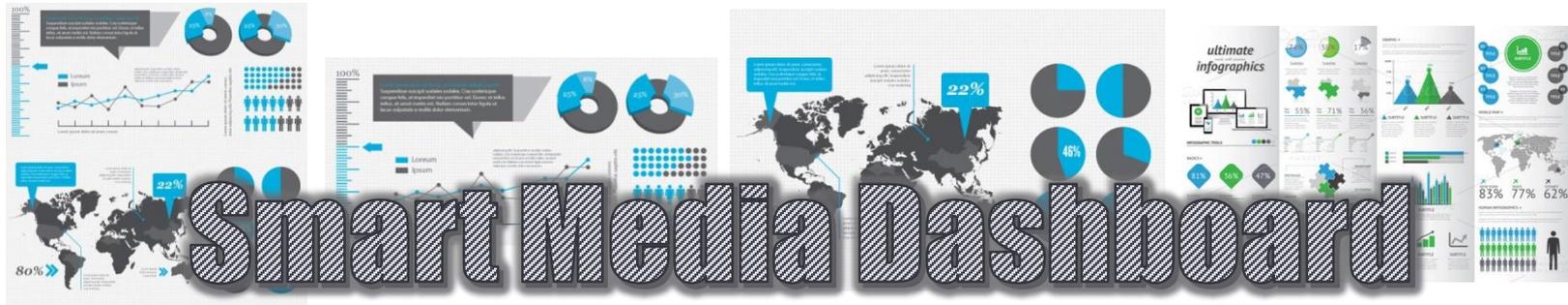
▶ 뉴스기사 지식베이스와 뉴스기사 분석서비스(가시화)의 완벽한 분리

- ▶ 아이돌 스타를 발굴하고 키워서(지식구축), 음악, 춤, 의상, 캐릭터 등을 부여하여 활동시킴으로써(지식분석서비스) 부가가치를 확보



뉴스기사 분석서비스 - 모듈화

▶ News Article Analysis Service



Visualization Area

Knowledge Area

“자동차 사고” - “방지하다” - “자율 응급 제동 시스템”
 “자율 응급 제동 시스템” - “동일하다” - “AEB”

→ 관계형 지식

(1)집터 고르기→(2)방위 정하기→(3)불 피우기→(4)주변 관찰하기

→ 절차적 지식

횡문근은 팔이나 다리 등 움직이는 부위에 붙어 있는 가로무늬 근육이다.

→ 정의형 지식

가장 앞선 기술력을 자랑하는 미국이 운용하는 군용 무인기 비율은 **2005년 5%에서 2010년 41%로 크게 늘어났다.**

→ 수치형 지식

뉴스기사 분석서비스 – Dashboard

▶ Smart Media Dashboard

- ▶ 뉴스기사 지식베이스 다차원 브라우저
 - ▶ 지식베이스에 저장된 다양한 종류의 지식을 실시간으로 표현



▶ **Smart Media Dashboard**

- ▶ 콘텐츠가 지식으로 변환되어 저장된 지식베이스를 활용하므로 서비스가 신속하게 제공될 수 있음 (속도가 빠름)
- ▶ 대시보드 템플릿 및 가시화 방법론의 확장을 통해 다양한 형태의 서비스 레이아웃 가능
- ▶ 개별 기사에 해당하는 가시화 요소는 서로 융합이 가능하여 새로운 Service Visualization 동적 생성이 가능함
 - ▶ Data Layer와 Visualization Layer가 완벽하게 분리됨

뉴스기사 분석서비스 - Dashboard

Compose message... 🔍 📧 🔒 ▶ Click to select a profile...

Home +

Scoop.it Dashboard

Scoop.it

Scoop'd by Buffy J. Hamilton
on eBooks, eReaders, and Libraries

Free Library of Philadelphia Resolves NOOK Accessibility Lawsuit - The Digital Shift
From thedigitalshift.com - Oct 30, 2012 09:19



Goldstein said that the current library e-reader landscape still poses significant accessibility problems.

Scoop'd by John Evans
on iPads in Education

Must Have Emergency Apps to Download on your iPad Before Hurricane Sandy Hits | PadGadget
From padgadget.com - Oct 30, 2012 08:54



Hurricane Sandy is due to hit land on late Monday or early Tuesday, and some areas are already experiencing severe weather conditions due to wind and rain.

Scoop'd by Ally Greer
on Content Marketing Info

How Hospitals Are Hustling as Content Creators

Search Scoop.it

Scoop.it

Scoop'd by Tocquigny Staff
on Tocquigny's Digital Marketing Daily

6 Dos & Don'ts of Content Marketing
From inc.com - Oct 30, 2012 02:01



Think having a bold, eye-catching homepage is enough? Think again. Outbrain's CEO shares his best strategies for attracting quality visitors to your site--and keeping them engaged.

Hey, your blog looks great! And those videos certainly add the finishing touch. Congratulations!

Now what?

Is it really enough to create your own content, push it out to your list, and wait for sales to boom?

This strategy may bring some measure of success to business owners with lists consisting of hundreds of thousands of contacts, but not for the average entrepreneur. According to Yaron Galai, the CEO of Outbrain, a content-discovery platform that places their clients' articles, mobile, and video content on high-profile publisher sites, you need to get creative about

Search Scoop.it

Scoop.it

Scoop'd by Robin Good
on Content Curation World

Social Content Curation: An Introductory Guide for Teachers and Students
From educause.edu - Oct 28, 2012 17:30



Robin Good: "7 Things You Should Know About Social Content Curation" is a technology brief from Educause which aims to introduce, explain and illustrate the emerging social curation trend and why it is relevant to teaching and learning.

From the official abstract: "An emerging class of online tools, including Pinterest, Scoop.it, EduClipper, and others, allows users to quickly and easily gather, organize, and share collections of online resources, particularly visual content.

These applications make it easy to collect and post disparate bits of content, providing visual groupings at a glance that can reveal important patterns.

In academic settings, they can facilitate more visual thinking and discussion among students while providing a means to share collections of online content."

Sent Tweets

(scoopit)

scoopit
2:22pm via Twitter for Mac
Hurricane #Sandy may be unprecedented in East Coast storm history via @APHumanGeog scoo.it/7qRiXF Stay safe, East Coast!
4 retweets

scoopit
2:01pm via Twitter for Mac
Have you found professional success on LinkedIn with the help of Scoop.it? Tweet us - we want to know!
3 retweets

scoopit
Oct 26, 4:52pm via Twitter for Mac
Thanks @wiredinstructor for sharing your story with us in this week's #ScoopitSpotlight!

scoopit
Oct 26, 4:50pm via Twitter for Mac
Dennis O'Connor's #Curation Surprise after Years of Knowledge and Content Discovery blog.scoop.it/en/2012/10/26/... #ScoopitSpotlight @wiredinstructor
1 retweets

gdecugis
Oct 26, 1:46pm via Scoop.it
+1! "Sites like Scoopit will give power to every day people" says @MarketingHits scoo.it/5BybWD
Retweeted by scoopit

briansolis
Oct 26, 11:12am via Tweet Button
We Need Social Producers: Catalysts for Conversations, Info & ROI | @scoopit scoo.it/5pkJsh
Retweeted by scoopit

scoopit
Oct 26, 12:29pm via Twitter for Mac
Tweetchat Reveals Secrets to Curation for Marketing blog.scoop.it/en/2012/10/26/... cc @MarketingHits #scoopitchat
1 retweets

scoopit
Oct 26, 10:50am via Twitter for Mac

content marketing

Search

alteryx
Oct 25, 2:17pm via Web
Our popular Presidential Election App predicts how your neighbors will vote. Find out how we do it, then give it a try: bit.ly/Td2yUq
Promoted by Alteryx

rachelloumiller
7:45pm via HootSuite
For Influential Content Marketing, Go Beyond Your Brand Fans ow.ly/eQR3C #scozbiz

nicomanesse
7:44pm via Tweet Button
¿Cómo el marketing 2.0 influye en el punto de venta? goo.gl/K7TcH

reverd_com
7:43pm via Tweet Button
7 Smart Content Marketing Ideas Anyone Can Use interactmedia.com/content-market... via @InteractMedia

mrs_cheeky
7:41pm via Twitter for iPad
RT @bryankramer Content Marketing: 7 Tips to Become the Signal Instead of the Noise goo.gl/T0pXN via @kathkruse

insight_corp_jp
7:39pm via web
RT @sunaomunakata: B2Bのコンテンツマーケティングに関する米国の最新レポート。成功している会社は、マーケティングの4割をコンテンツマーケティングに投じている！ ow.ly/eRW3x

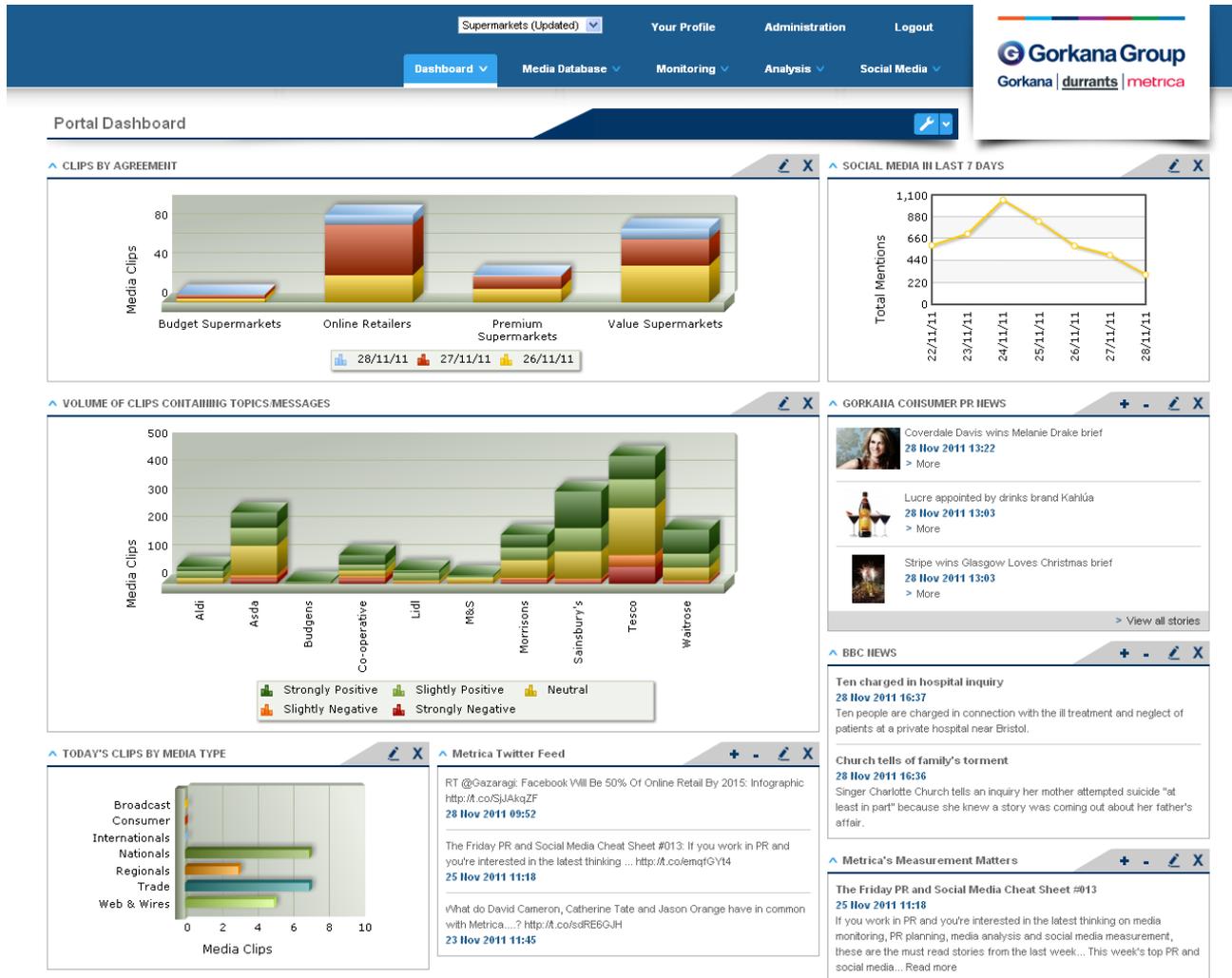
Pentavere
7:36pm via web
Great examples of forward thinking hospitals engaging patients and delivering greater value. business2community.com/content-market... #

sunaomunakata
7:36pm via HootSuite
B2Bのコンテンツマーケティングに関する米国の最新レポート。成功している会社は、マーケティングの4割をコンテンツマーケティングに投じている！ ow.ly/eRW3x

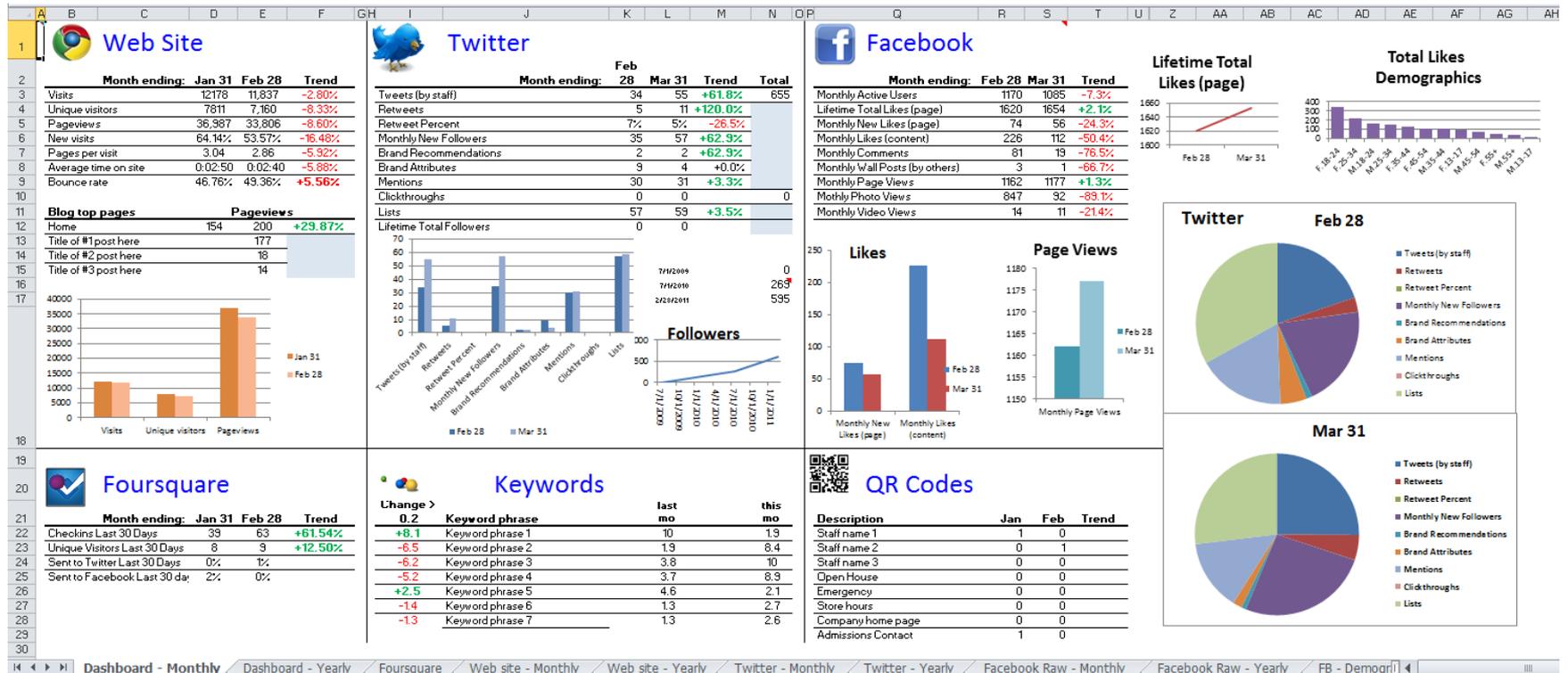
뉴스기사 분석서비스 - Dashboard



뉴스기사 분석서비스 – Dashboard



뉴스기사 분석서비스 - Dashboard



뉴스기사 분석서비스 – Dashboard



VISIBLE INTELLIGENCE

Dashboard Search Monitor Analyze Engage

Technology Home Help & Support User Name Sign Out

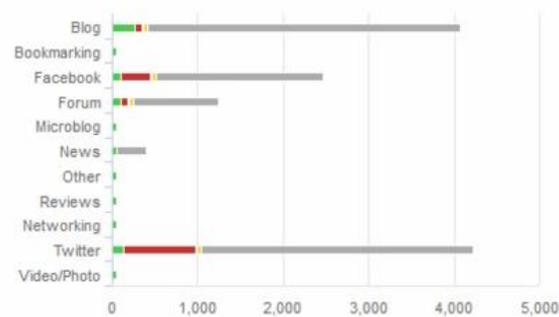
Sprint Brand Volume



Sprint Brand

[View Content Volume in Monitor](#)

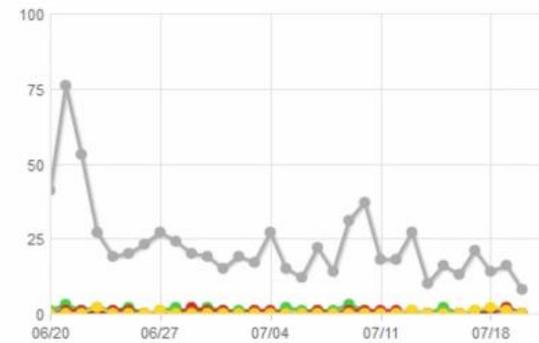
Sprint Brand Sentiment by Media Type



Positive Negative Mixed Neutral

[View Content Media Type Sentiment in Monitor](#)

Sprint Windows Phone 7 Sentiment



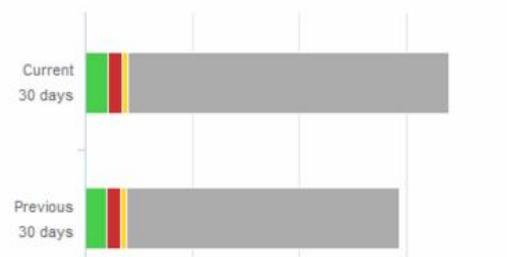
Positive Negative Mixed Neutral

[View Content Sentiment in Monitor](#)

Sprint Customer Service

Sprint customer service, tech support & their stores are not in sync, whi
 Sprint Nextel(S): Good Customer Service, Network Reliability And Com
 Sprint Nextel(S): Good Customer Service, Network Reliability And Com
 KTK 98.5 » Are you getting sales calls and spam on your cell phone?
 @LaLaCocaColas @realrodjohnson @empathyone it would be a good
 Supervisor Customer Service Technical Support at Sprint (Englewood, t
 Dear Sprint Customer Service, *I officially HATE YOU! How the hell are u

AT&T Network Image Sentiment



Verizon Network Image Sentiment by Media Type



추진 전략 - 뉴스기사 지식베이스

▶ 두 가지 해결해야 할 중요한 과제

▶ 지식 표현 방법

- ▶ 관계형, 절차적, 정의형, 수치적 지식을 어떻게 저장해야 하나?
- ▶ 기존에 연구된 방법론이 존재하는지 조사 필요
- ▶ 다양한 지식을 포괄할 수 있는 확장성 있는 지식 표현 방법 필요
 - ▣ 향후에 새로운 형태의 지식이 도출될 가능성이 높음

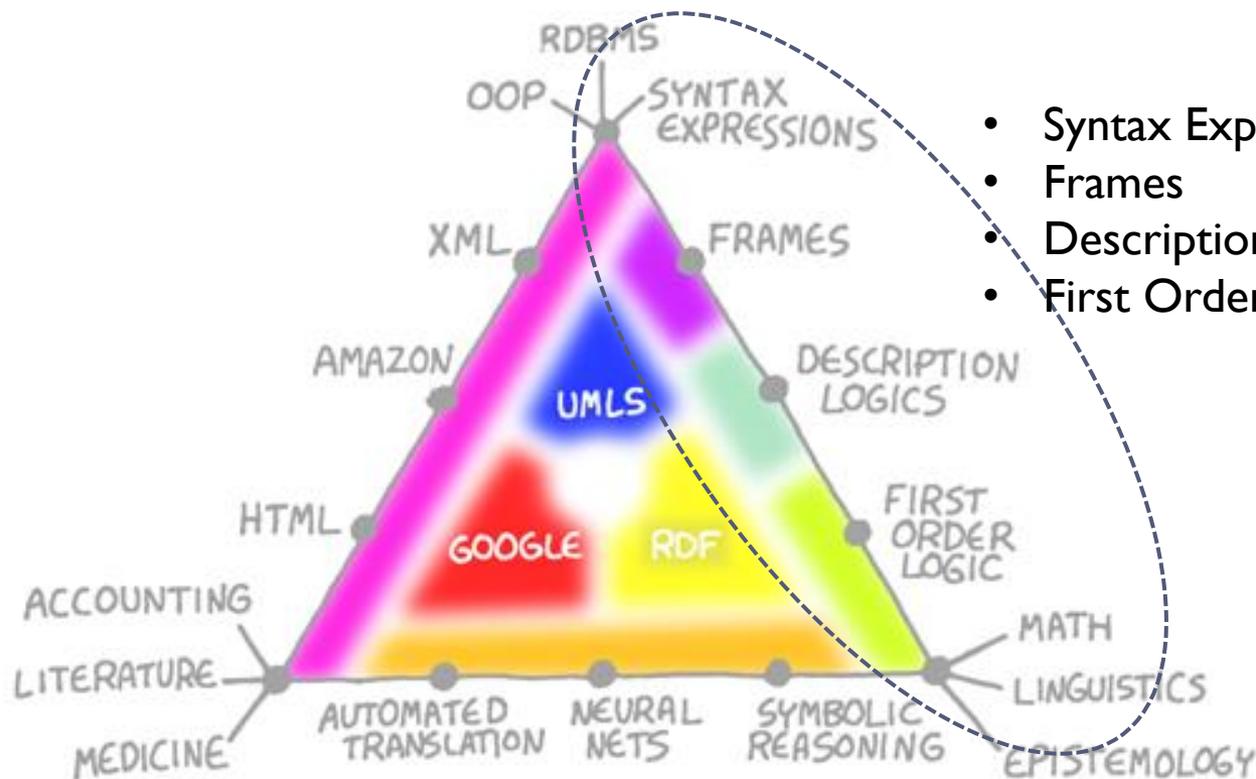
▶ 지식 구축 방법

- ▶ 자동화가 불가능한 상황에서 가장 효율적인 반자동 지식 구축 프로세스 도출 필요
- ▶ 장기적으로 자동화를 추구하기 위한 기반 엔진의 성능 개선 방안 수립

추진 전략 - 뉴스기사 지식베이스

▶ 지식 표현 방법

- ▶ 다양한 지식 표현 방법 존재



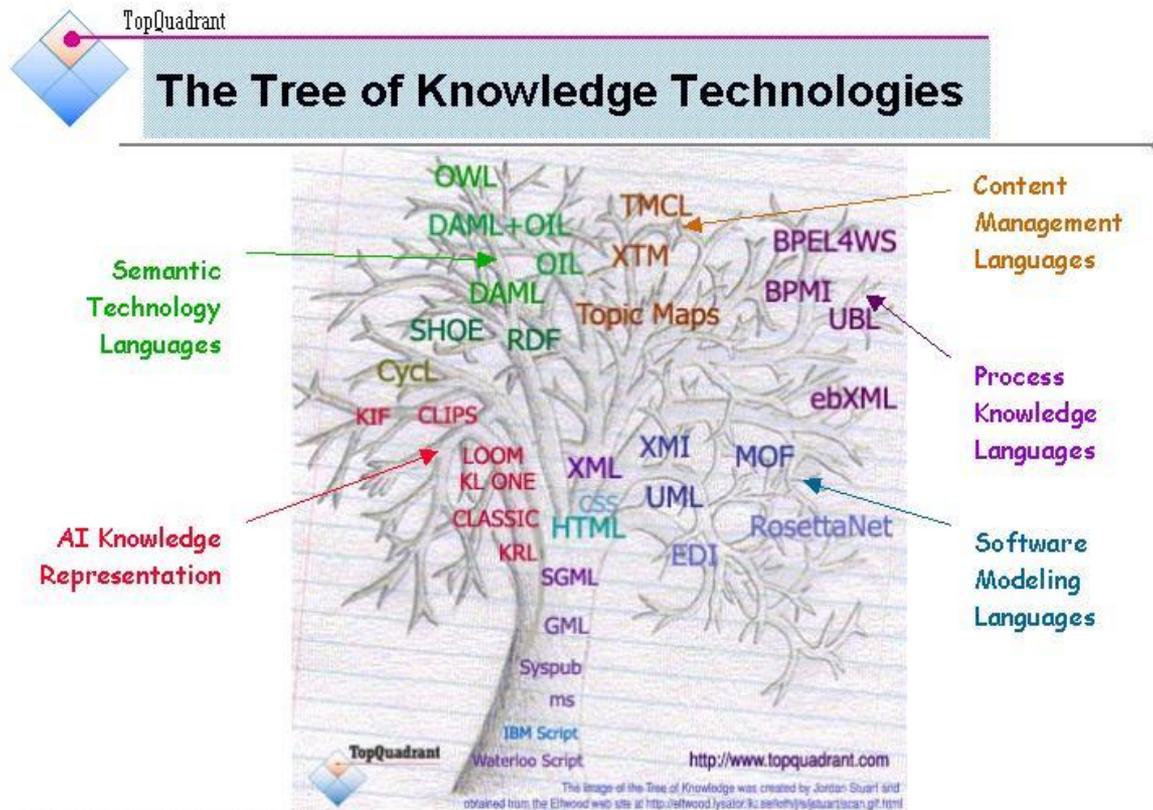
- Syntax Expressions
- Frames
- Description Logics
- First Order Logic

<http://www.lisperati.com/tellstuff/>

추진 전략 - 뉴스기사 지식베이스

▶ 지식 표현 방법

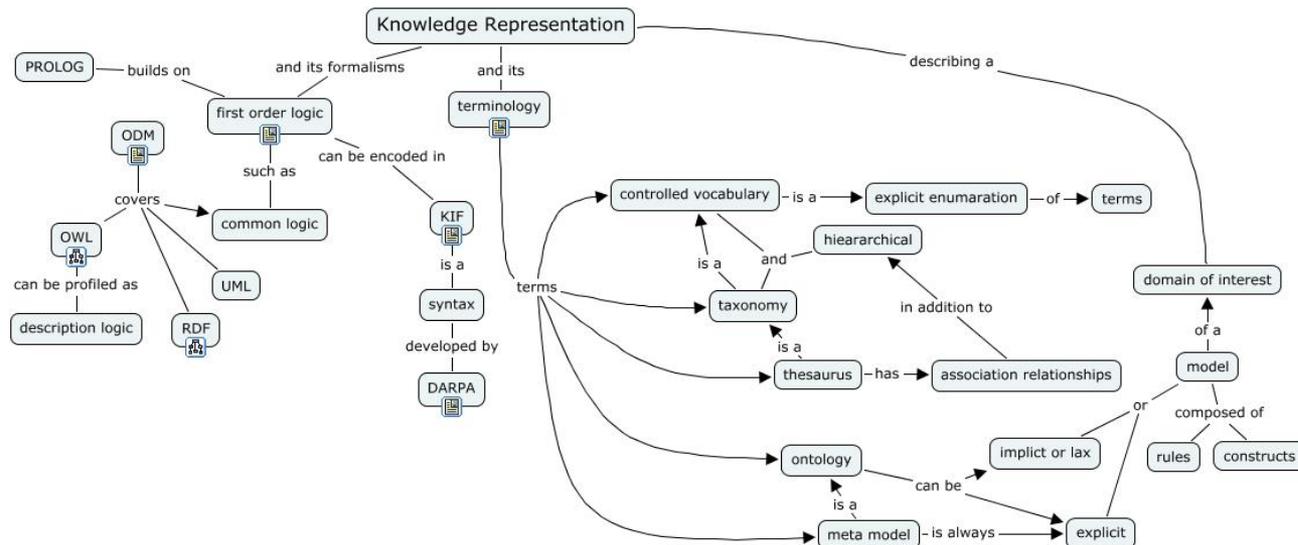
- ▶ 다양한 지식 표현 방법 존재



추진 전략 - 뉴스기사 지식베이스

▶ 지식 표현 방법

- ▶ 목표 지향적 지식 표현 방안 마련
 - ▶ 최종 목표는 뉴스기사 콘텐츠의 가시화이므로 이에 표현되어야 할 지식을 잘 표현할 수 있는 스키마 고안이 필요함
 - ▶ 인공지능이나 시맨틱 웹에서 이야기하는 거창한 추론이나 지식 확장은 필요가 없음
 - ▶ 콘텐츠 사용자에게 필요한 지식에 대해서만 집중



추진 전략 - 뉴스기사 지식베이스

▶ 지식 표현 방법

- ▶ 확장이 용이하도록 지식 표현 스키마 구성
 - ▶ 관계형 지식 → RDF (Subject-Predicate-Object) (RDB도 가능)
 - ▶ 절차적 지식 → RDB (관계형 데이터베이스에 저장 가능)
 - ▶ 정의형 지식 → RDB (관계형 데이터베이스에 저장 가능)
 - ▶ 수치형 지식 → RDB (관계형 데이터베이스에 저장 가능)

- ▶ 세부적인 방법론에 대해서는 추후 연구가 필요함
 - ▶ 위에서 기술한 지식 외에도 다양한 지식이 존재할 수 있으므로 이를 쉽게 저장하기 위한 스키마를 고안해야 함
 - ▶ 지나치게 학술적으로 복잡하게 구성하는 것은 위험함

추진 전략 - 뉴스기사 지식베이스

▶ 지식 구축 방법

- ▶ 초기에는 견고한 수동 프로세스를 구동함으로써 즉시성 있는 가시적 성과는 물론 높은 수준의 저널리즘 지식 구축 인력 (Journalism Knowledge Engineers) 양성 도모
 - ▶ 수작업 대비 자동화 비율 → 90:10

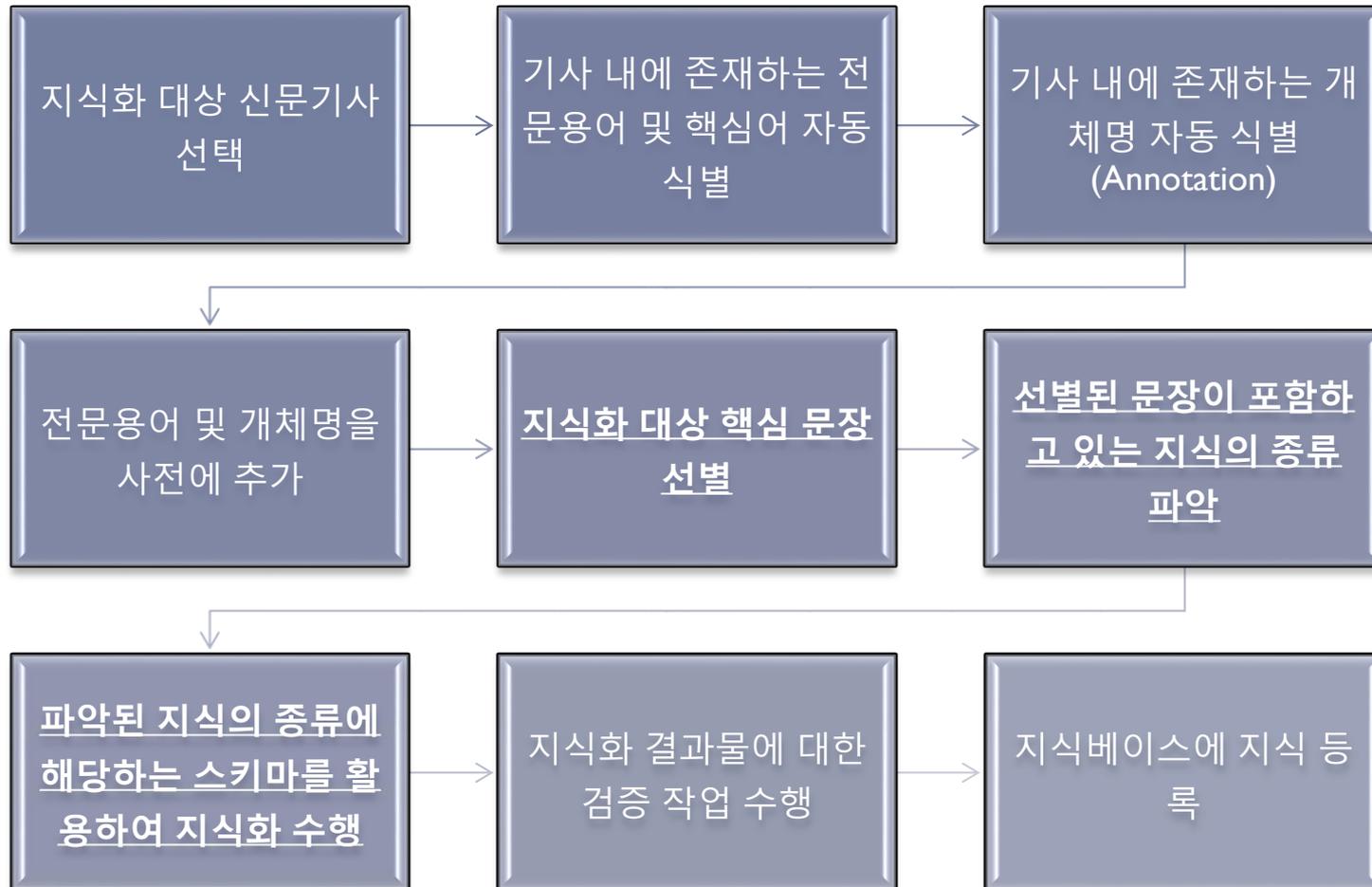
- ▶ 지식 구축 과정에서 도출되는 지식베이스, 사전, 온톨로지 등을 기반으로 자동화 엔진 개발
 - ▶ 개발된 엔진 성능 개선을 위한 지속적 튜닝 노력 경주

- ▶ 성능이 안정화된 엔진을 기반으로 자동화 수준 향상
 - ▶ 수작업 대비 자동화 비율 → 40:60



추진 전략 - 뉴스기사 지식베이스

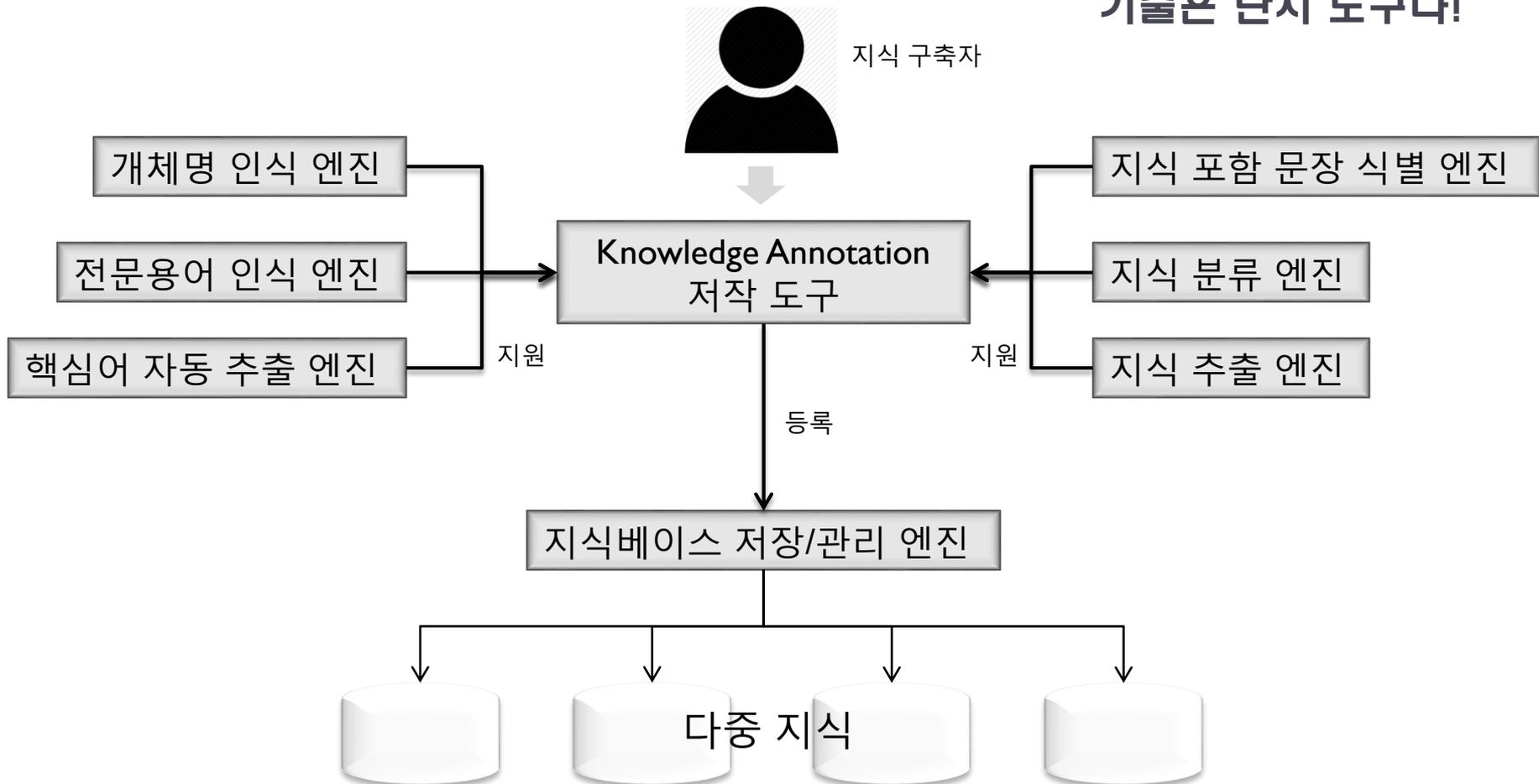
▶ 지식 구축 방법 (프로세스)



추진 전략 - 뉴스기사 지식베이스

▶ 지식 구축 방법 (시스템 아키텍처)

기술은 단지 도구다!



추진 전략 - 뉴스기사 지식베이스

▶ 지식 구축 방법 (인력)

▶ 지식 구축자

▶ 조건

- 데이터베이스 및 스키마에 대한 일반적인 이해
- 언어 분석 감각 보유
- 데이터베이스 구축 경험

▶ 교육

- 초기에는 전문가의 지도아래 교육 필요
- 기사 및 칼럼에서 핵심적인 문장 및 문구를 발견하고 이를 구조화할 수 있는 감각에 대한 노하우 필요

▶ 효과

- 위와 같은 능력과 경험을 가진 지식 공학 인력들을 많이 보유한다면 향후 신문기사 외의 다양한 콘텐츠에 대한 지식화가 용이함
- 새로운 창조적 일자리 창출

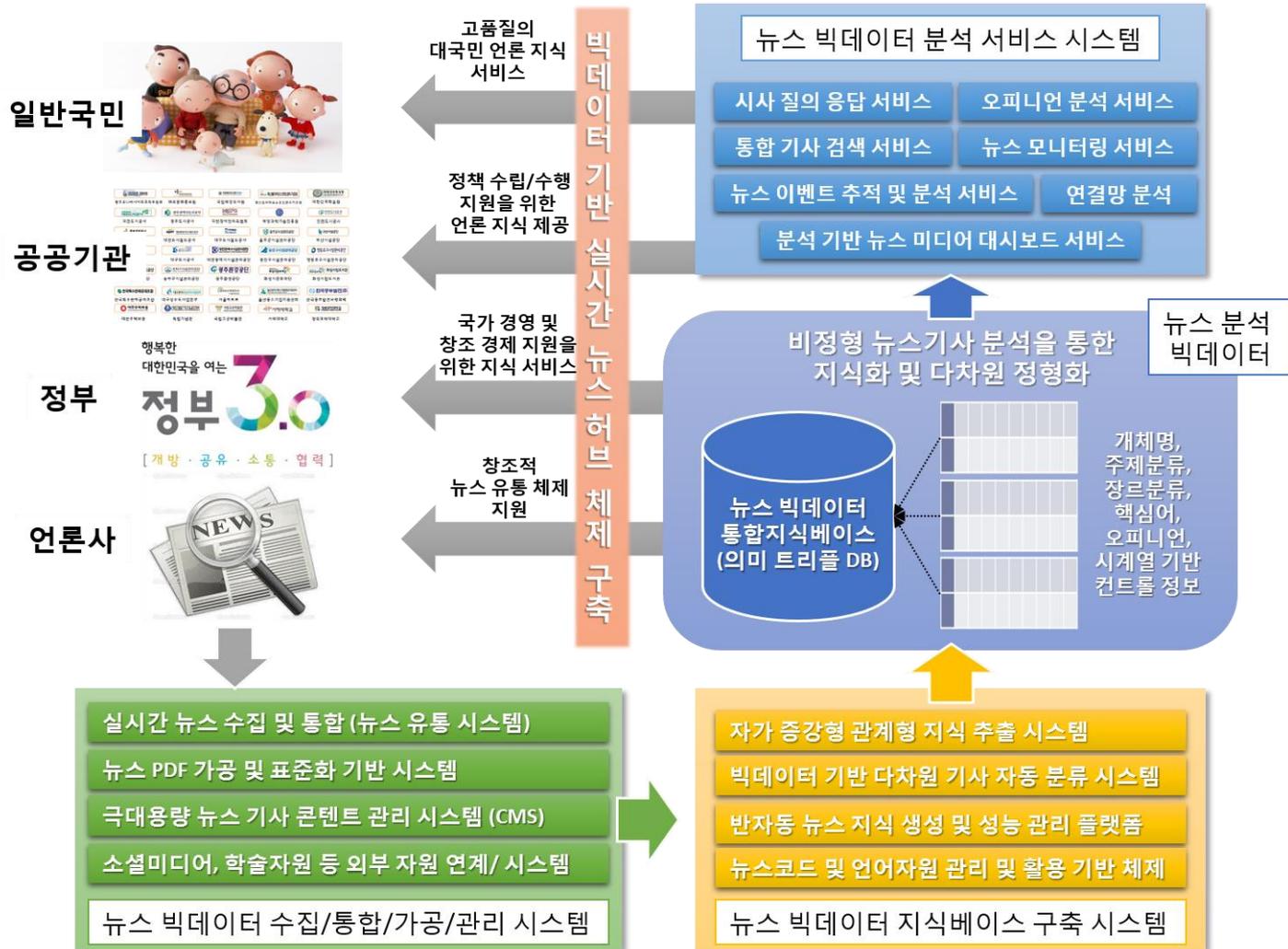
추진 전략 - 기술적 요구사항 분석

- ▶ **지식베이스 반자동 구축에 필요한 기술**
 - ▶ 형태소 분석 기술
 - ▶ 신문기사 자동 분류 기술 (주제, 장르, 인물, 지역, ...)
 - ▶ 핵심어 추출 기술
 - ▶ 개체명 (인명, 지명, 상호명) 인식 및 식별 기술
 - ▶ 개체명 간 관계 추출 기술
 - ▶ 통합 지식베이스 구축 지원 도구
- ▶ **스마트 미디어 대시보드 서비스 구축에 필요한 기술**
 - ▶ 데이터베이스 연계 기술
 - ▶ 추론 기술
 - ▶ 인터페이스 뷰 자동 생성 기술 (Authoring Tool)

추진 전략 - 기술적 요구사항 분석

- ▶ **지식베이스 반자동 구축에 필요한 기술**
 - ▶ 형태소 분석 기술 (상)
 - ▶ 신문기사 자동분류 기술 (중)
 - ▶ 핵심어 추출 기술 (하)
 - ▶ 개체명 인식 및 식별 기술 (중)
 - ▶ 개체명 간 관계 추출 기술 (하)
 - ▶ 통합 지식베이스 구축 지원 도구 (상)
- ▶ **스마트 미디어 대시보드 서비스 구축에 필요한 기술**
 - ▶ 데이터베이스 연계 기술 (상)
 - ▶ 추론 기술 (중상)
 - ▶ 인터페이스 뷰 자동 생성 기술 (Authoring Tool) (상)

시스템 구성도



통합 기사 지식베이스 구축 및 분석형 대시보드 서비스 플랫폼 기술 개발

언론정보에 최적화된 요소 기술 개발 및 적용
통합 지원 도구 개발을 통한 반자동 프로세스 편의성 극대화

- 뉴스기사 전용 개체명 인식기 개발
- 뉴스기사 자동 분류 엔진 개발
- 뉴스기사 내 핵심어 추출엔진 개발
- 엔진 최적화 통합 도구 개발

- 개체명 인식 후처리 정제 도구 개발
- 자동 분류 검증 및 수정 도구 개발
- 지식베이스 구축 워크플로우 플랫폼 개발

지식베이스 구축을 위한
핵심 요소 엔진 개발

기사 지식베이스 구축
지원 도구 개발

분석형 가시화 요소
자동 생성 플랫폼 개발

News KB 구축/활용 및
지속적 성능 향상 지원

- 상황별/유형별 분석형 가시화 요소 템플릿 DB 구축 및 분류
- 템플릿 기반 가시화 요소 반자동 저작 지원 도구 개발
- KB 연계 및 추론 기술 개발

- News KB 구축 워크플로우 관리 및 통합 지원 체제 구축
- News KB 기반 응용서비스 개발 및 적용

뉴스기사 지식화를 위한 인프라 체제 구축 지원

- 대용량 뉴스기사에 대한 디지털화 및 데이터베이스 구조화
- News KB 기반 웹 서비스 및 분석형 대시보드 서비스 생성 관련 요구사항 제공

추진체계 (KB & Smart Media Dashboard)

통합 기사 지식베이스 구축 및 스마트 미디어 대시보드 서비스 플랫폼 개발

- 지식베이스 구축을 위한 개체 명인식, 분류, 핵심어 추출 엔진 개발
- 지식베이스 구축 지원 도구 개발
- 지식베이스 워크플로우 구축

기반 구축

- 지식베이스 구축을 위한 핵심 엔진 성능 고도화
- 지식베이스 기반 기사 분석 및 응용 시스템 개발
- 워크플로우 관리 및 최적화

고도화 및 활용

- 분석형 미디어 대시보드 반 자동 생성 저작 도구 개발
- 지식베이스 연계 추론 기술 개발 및 적용
- News KB 운영 및 확산

본격 운영/확산

지식베이스 구축의 효율화

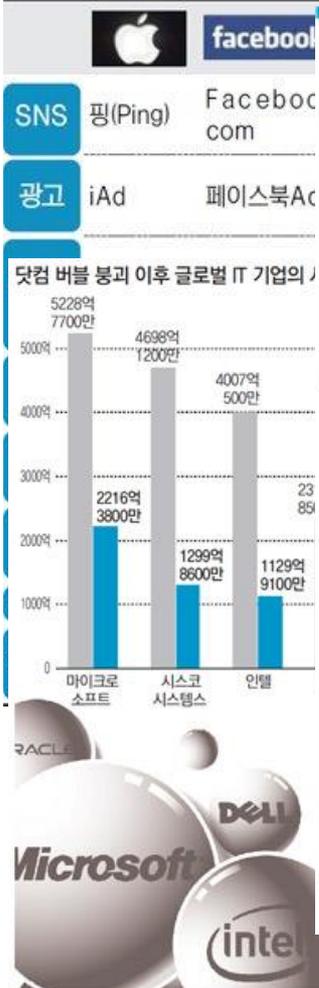
수준 높은 기반 핵심 엔진 개발 및 최적화로 신문기사 지식베이스 구축의 효율성 확보

지식베이스 활용의 극대화

웹 기반 응용 서비스 및 스마트 미디어 대시보드 서비스 플랫폼 개발 및 활용을 통해 지식베이스의 확산 도모

결론 - 빅데이터 시대의 성공 불문율

미국 5대 IT기업 전면전 현황



구글의 혁신 생태계

