

Opinion Mining, 정성과 정량분석

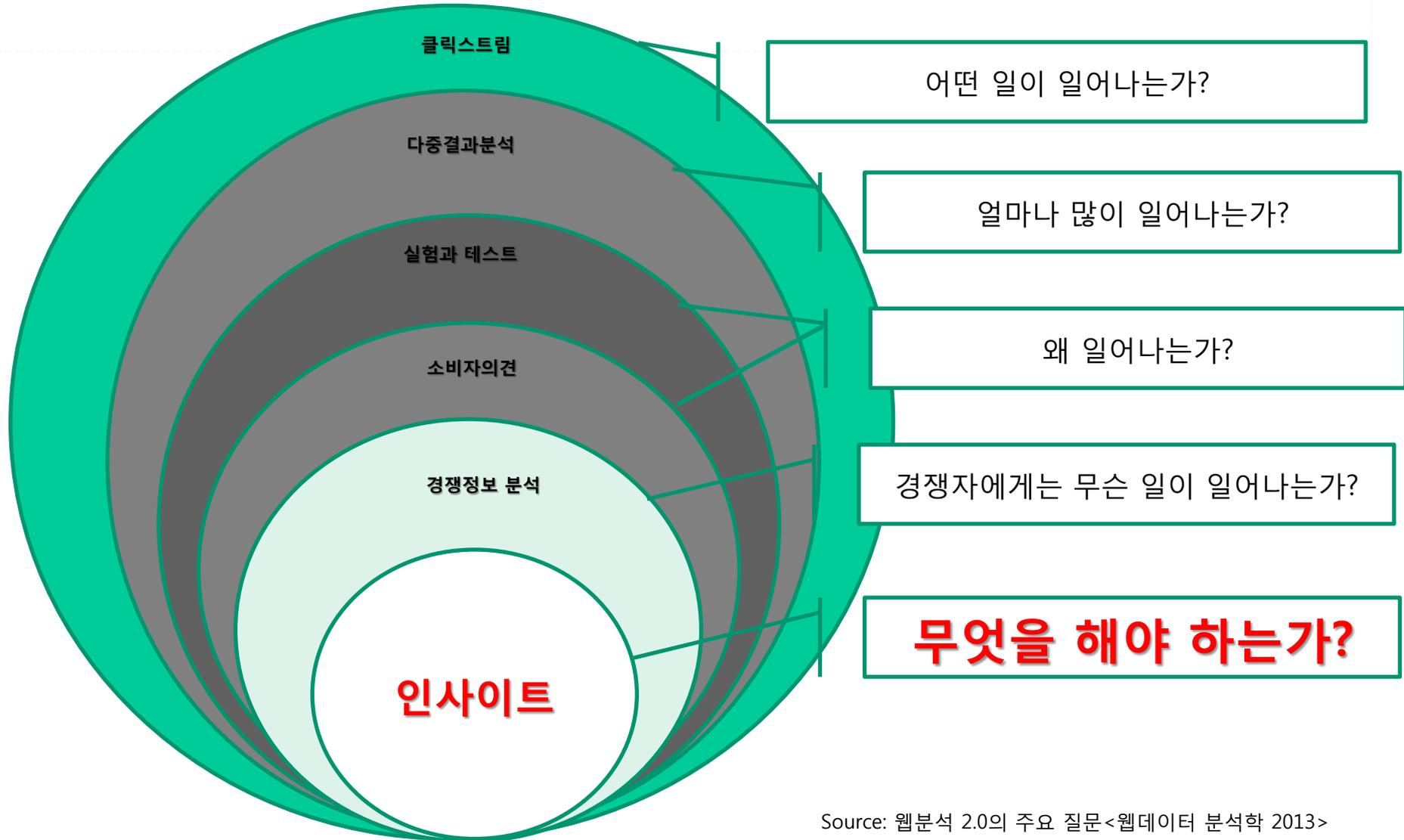
Getting Smart Insight from Text Big Data

2014.10.08



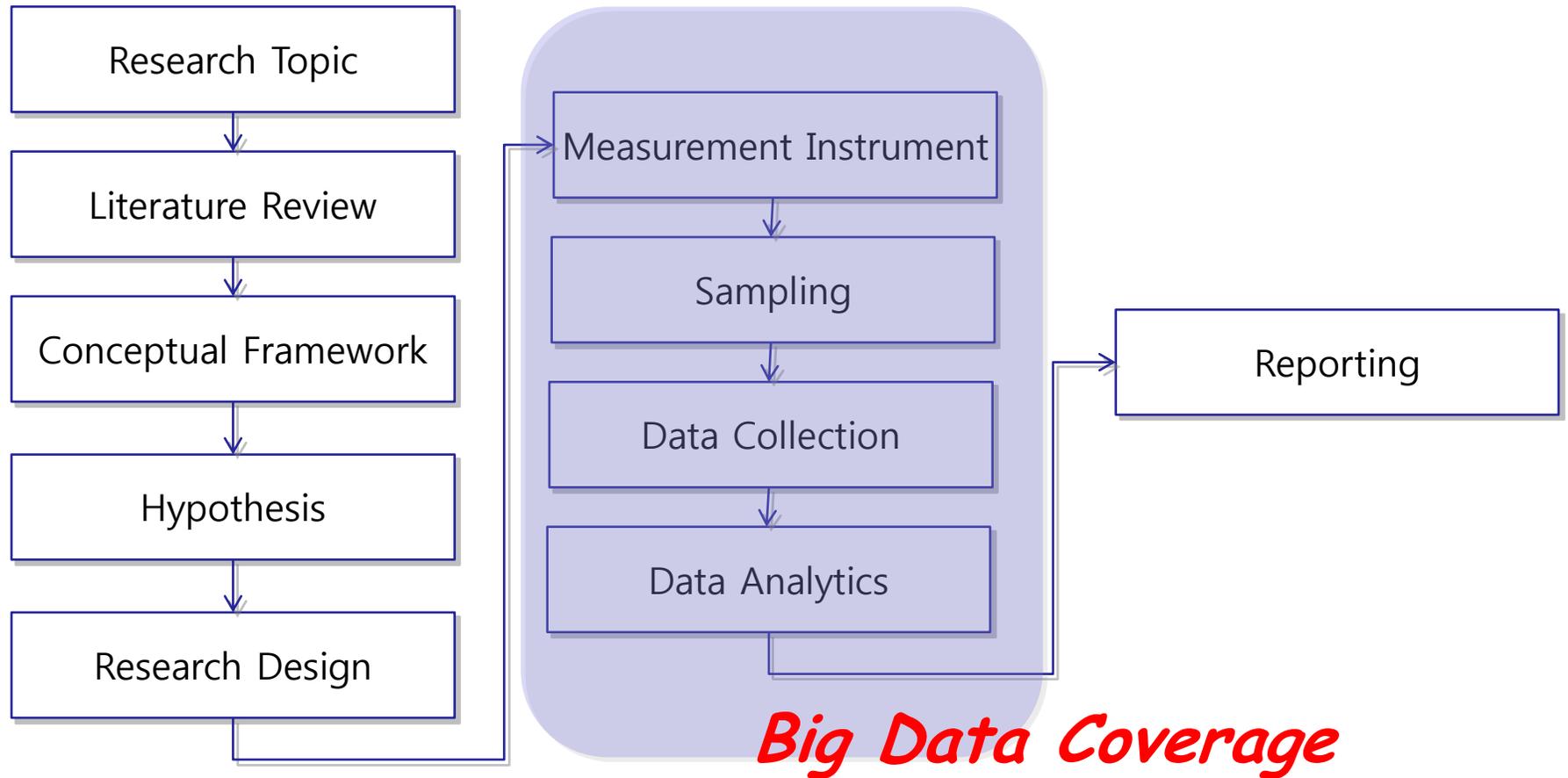
Data Scientist

Web Data 분석



Source: 웹분석 2.0의 주요 질문 <웹데이터 분석학 2013>

Research Process



자연언어처리(NLP)

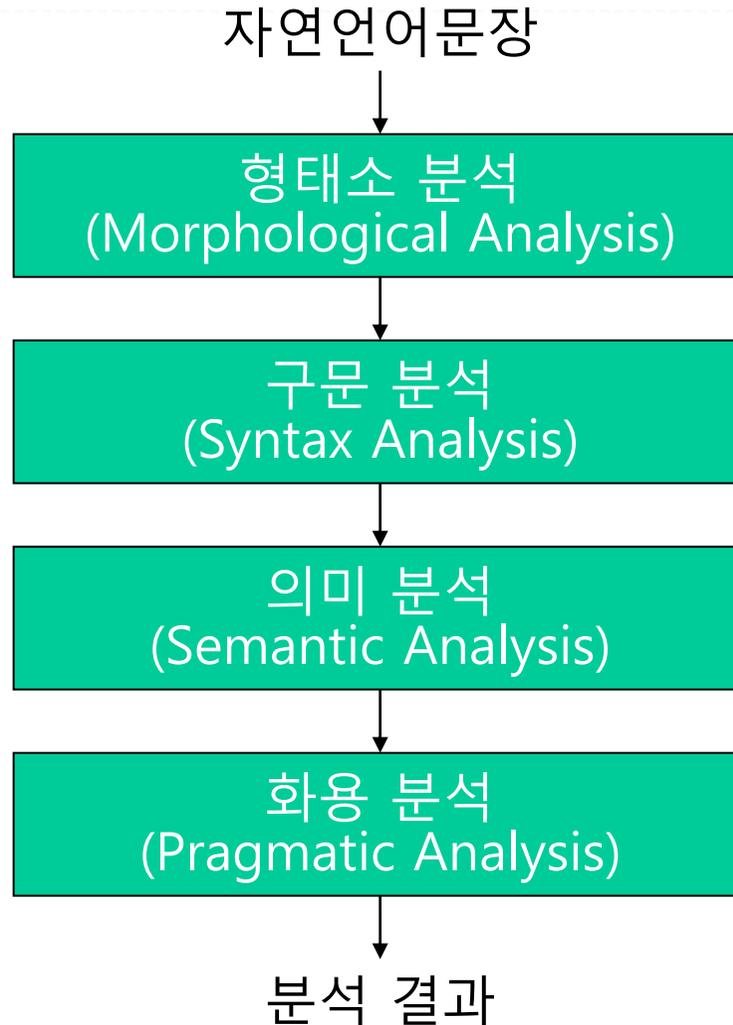
□ 자연언어처리란?

- 컴퓨터를 통하여 인간의 언어를 처리하고 이용하려는 학문 분야
- 인간의 언어를 이해하고, 이를 바탕으로 각종 정보처리에 적용함으로써 보다 빠르고 편리한 정보 획득을 목표로 함

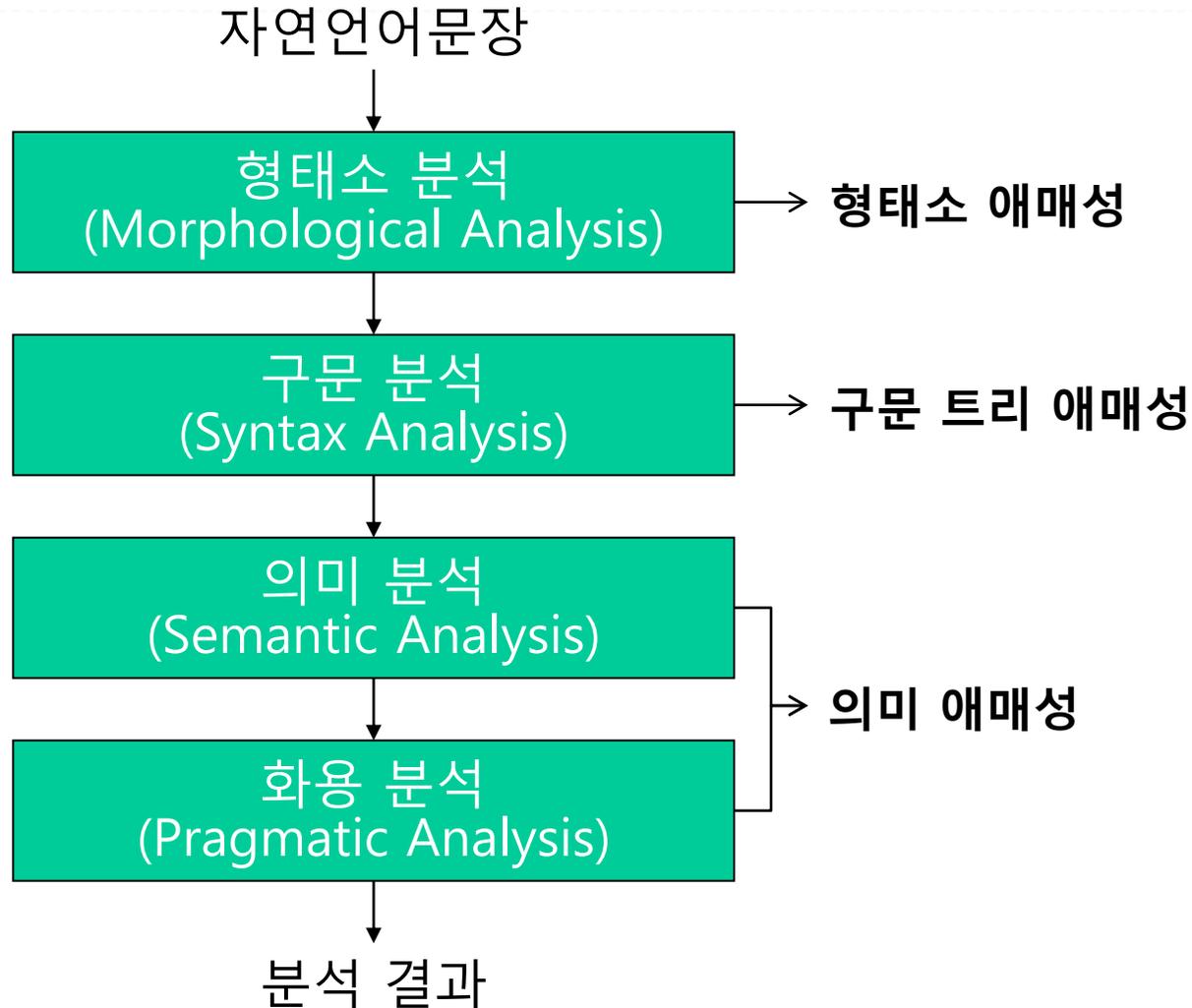
□ 자연언어처리 응용 분야

- 인간의 언어가 사용되는 실세계의 모든 영역
- 정보검색, 기계번역, 자동통역, 문서작성, 문서요약, 문서 분류, 철자 오류 검색 및 수정, 문법 오류 검사 및 수정, 음성합성, 음성인식

자연언어 분석 단계: levels of NLP

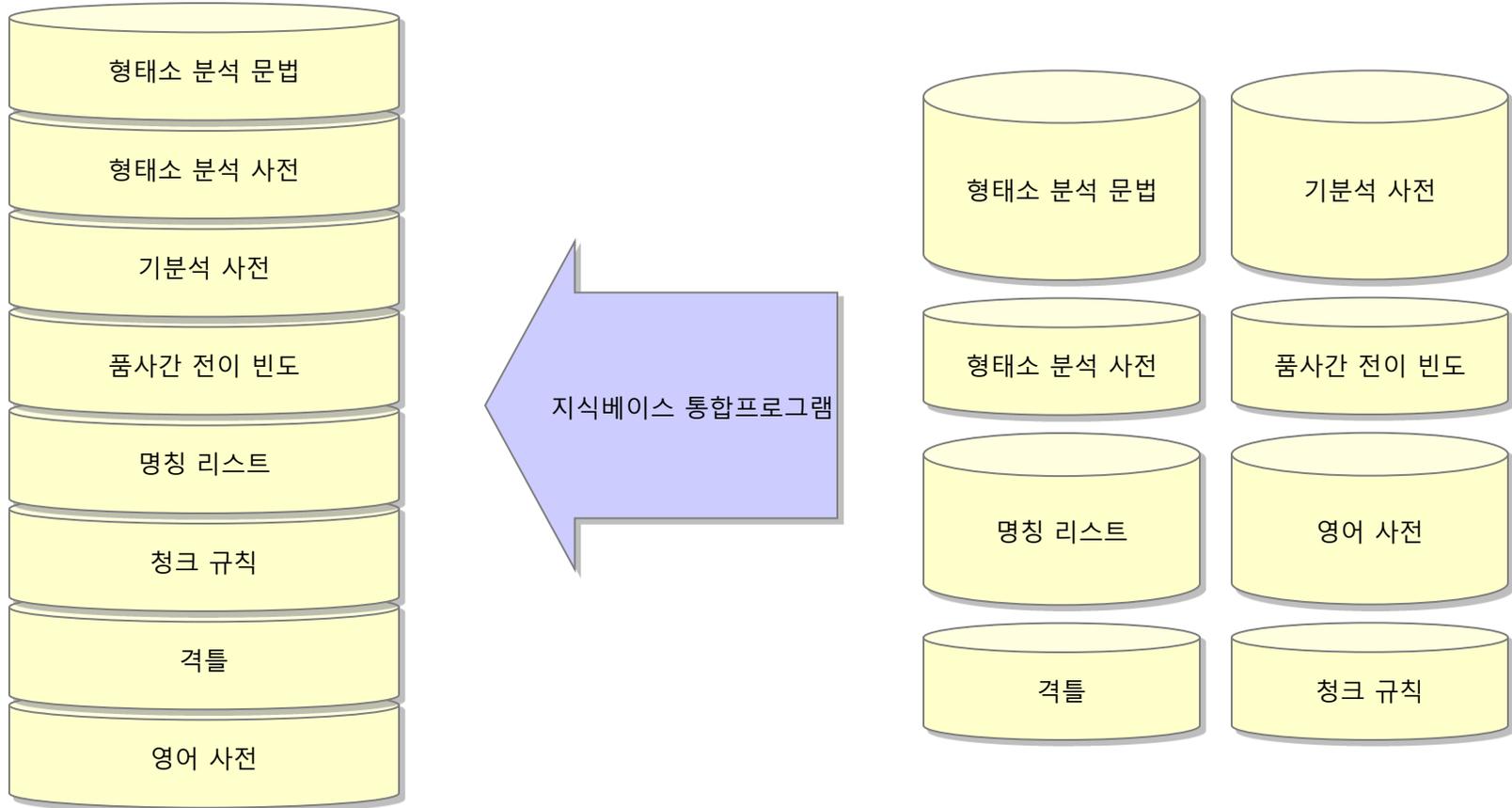


통계적 처리 과정의 필요성



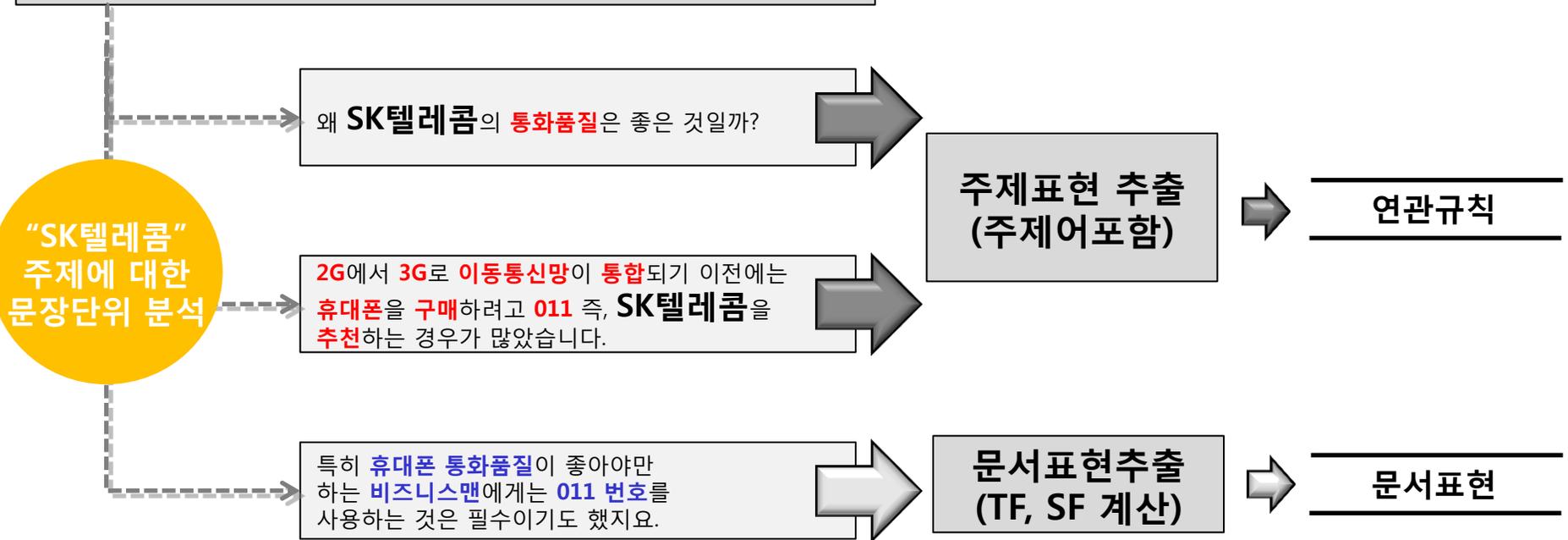
통합 지식베이스에는 여러가지 언어처리를 위한 지식이 통합되어 있다.

□ 사용하는 지식베이스

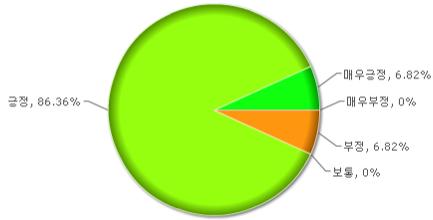


문서표현과 주제표현의 비교를 통해 문서의 주제적합여부를 판단함.

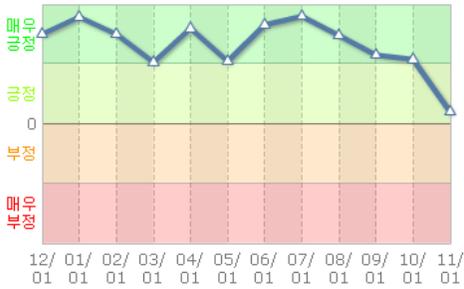
왜 **SK텔레콤**의 **통화품질**은 좋은 것일까?
2G에서 **3G**로 **이동통신망**이 **통합**되기 이전에는 **휴대폰**을 **구매**하려고 **011**
즉, **SK텔레콤**을 **추천**하는 경우가 많았습니다. 특히 **휴대폰 통화품질**이 좋아야만
하는 **비즈니스맨**에게는 **011 번호**를 사용하는 것은 필수이기도 했지요.



전체 문서 감성 정보



Opinion 리포트에서 확인 가능한 정보를 아이콘으로 표시



【공부정도 평가값 변화】

공부정도 변화 정보

주요 표현 정보

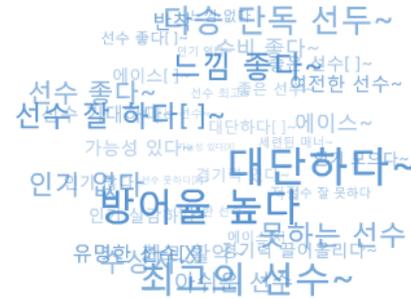
【주요표현 TOP10】 (분석일 : 년간)

순위	표현	추출수
1	에이스	232
2	수상자	35
3	다승 단독 선두	14
4	가능성 있다	9
5	대단하다	8
6	대단하다(+)	8
7	경기력 있다	8
8	최고 활약	8
9	최고의 선수	6
10	인기 실감하다	5

【표현볼륨 및 공부정 비율 변화】

공부정도 비율 변화 정보

주요 표현 구형 tag cloud 표현



주요 표현 tag cloud로 표현

【주요표현 Tagcloud - 공부정순】 (분석일 : 년간)

볼륨순

선수 잘 하다[+]: 2 | 좋은 선수[+]: 2 | 경기력 끌어올리다: 3 | 경기력 있다[+]: 1 | 대단한 선수: 1 | 인기 많다: 2 | 선수 최고: 2 | 선수 좋다[+]: 1 | 인기 일다: 2 | 인기 모으다: 2 | 에이스: 232 | 경기력 선보이다: 1 | 빛나는 선수: 2 | 최고의 선수: 6 | 대단하다: 8 | 선수 대단하다: 1 | 경기력 있다: 8 | 에이스[+]: 5 | 대단하다[+]: 8 | 인기 실감하다: 5 | 수상자: 35 | 다승 단독 선두: 14 | 가능성 있다: 9 | 한국 최고의 선수: 1 | 선수 못하다[+]: 2 | 선수 기대되다: 1 | 선수 별로이다[+]: 1 | 선수 권장다: 1 | 눈에 띄는 선수: 1 | 방어를 낮다: 1 | 가능성 있다[-]: 1 | 평균구속 상승: 1 | 여전한 선수: 2 | 선수 좋다: 3 | 세련된 매너: 2 | 방어를 좋다: 1 | 유명한 선수: 2 | 수비 좋다: 5 | 선수 잘 하다: 1 | 느낌 좋다: 2 | 좋은 선수: 3 | 최고 활약: 8 | 선수 기대하다: 3 | 선수 한국 최고: 1 | 선수 좋지 않다[+]: 1 | 에이스[+]: 2 | 방어를 좋다[R+]: 1 | 저공: 3 | 반칙: 2 | 선수 못하다: 1 | 떨어지는 선수: 1 | 나이 막다: 1 | 가능성 있다[+]: 2 | 방어를 높다: 3 | 느낌 없다: 2 | 인기 없다: 3 | 못하는 선수: 3 | 아쉬운 선수: 4 | 선수 잘 못하다: 1 | 유명한 선수[+]: 1 | 뻔한 선수: 1 | 좋지 않은 선수: 1 | 안타까운 선수: 1 | 문제있는 선수: 1 | 못한 느낌: 1



【 출처(호스트)별 긍부정도 및 추출표현볼륨 】

출처(호스트)별 긍부정도 및 표현볼륨 비교



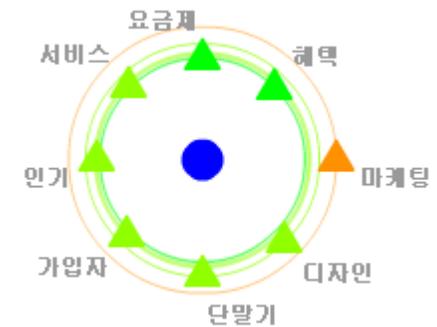
【 채널별 긍부정도 및 추출표현볼륨 】

채널별 긍부정도 및 표현볼륨 비교

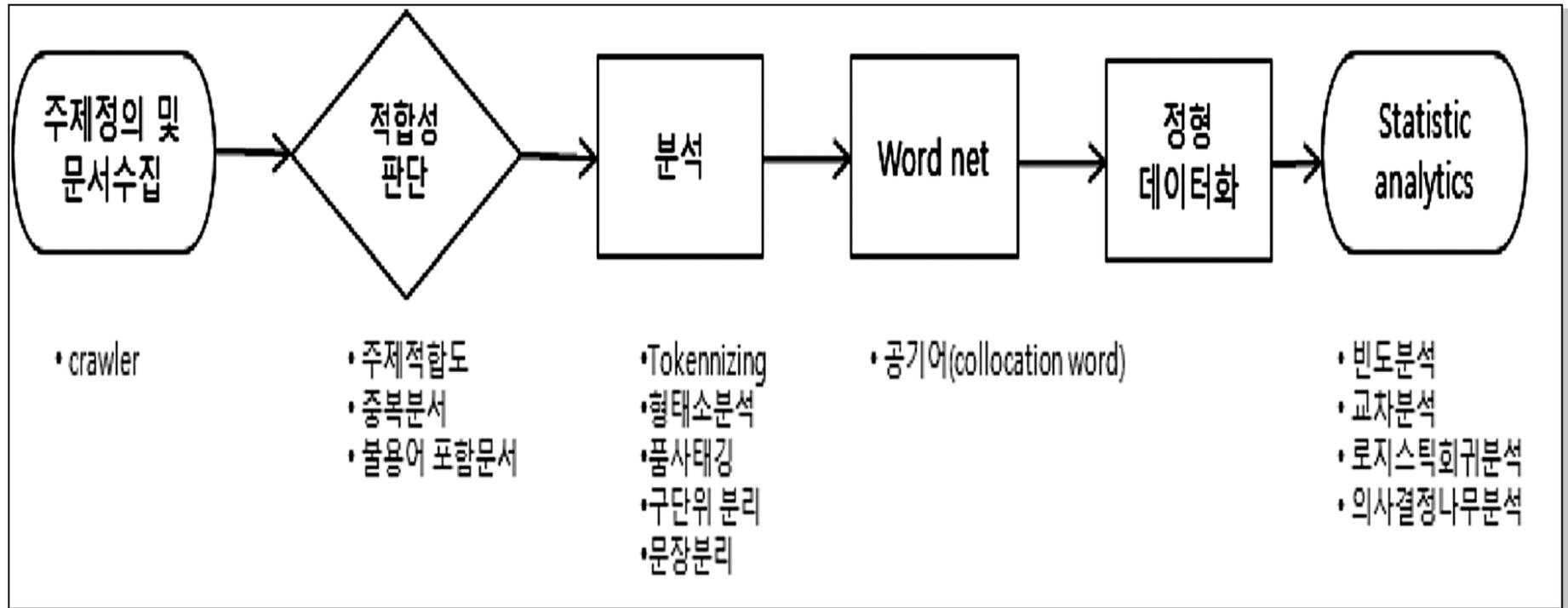
속성/출처별 표현볼륨



속성 포지션 맵



정형화된 Text Big Data 분석의 Approach



분석지의 확장을 위한 소셜 빅데이터 활용연구 - 국내 '빅데이터' 수요공급 예측 - *

a Study on Using Social Big Data for Expanding Analytical Knowledge
- Domestic Big Data supply-demand expectation -

김 정 선 (Jung-Sun Kim)

이화여자대학교 대학원 디지털미디어학부 (smileaga.in222@ewhain.net)

권 은 주 (Eun-Ju Kwon)

이화여자대학교 대학원 디지털미디어학부 (ej.kwon@ewhain.net)

송 태 민 (Tae-Min Song)**

한국보건사회연구원 통계정보연구실장 (tmsong@khisara.re.kr)

ABSTRACT

Big data seems to change knowledge management system and method of enterprises to large extent. Further, the type of method for utilization of unstructured data including image, video, sensor data and text may determine the decision on expansion of knowledge management of the enterprise or government. This paper, in this light, attempts to figure out the prediction model of demands and supply for big data market of Korea through data mining decision making tree by utilizing text bit data generated for 3 years on web and SNS for expansion of form for knowledge management. The results indicate that the market focused on H/W and storage leading by the government is big data market of Korea. Further, the demanders of big data have been found to put important on attribute factors including interest, quickness and economics. Meanwhile, innovation and growth have been found to be the attribute factors onto which the supplier puts importance. The results of this research show that the factors affect acceptance of big data technology differ for supplier and demander. This article may provide basic method for study on expansion of analysis form of enterprise and connection with its management activities.

Keywords: knowledge Management, Big Data, Text Mining, Data Mining, Supply-Demand Prediction

1. 서론

최근 스마트폰, 스마트TV, RFID, 센서 등의 급속한

보급 및 모바일 인터넷과 소셜미디어의 확산으로 데이터량이 기하급수적으로 증가하고 데이터의 생산, 유통, 소비 체계에 큰 변화를 주면서 데이터가 경제적 자산이 될 수 있는 빅데이터 시대를 맞이하게 되었다[송태민, 2012]. 세계 각국의 정부와 기업들은 빅데이터가 향후 국가와 기업의 성패를 가름할 새로운 경제적 가

* 논문접수일 : 2014년 7월 9일; 1차수정일: 2014년 8월 1일; 2차수정일: 2014년 9월 4일; 게재확정일 : 2014년 9월 12일

** 교신저자

- 2011년 1월 ~ 2013년 12월
- 뉴스, Blog, 카페, 게시판, SNS...
- 빅데이터, Big Data
(excluded 컨퍼런스, 전시회)
- 총 70,944건

5D 1C

- *Data Research Design*
- *Data Quality Assurance*
- *Data Organization*
- *Data verification*
- *Domain Knowledge*
- *Creativity*



감사합니다.

Jungsun.kim@sk.com