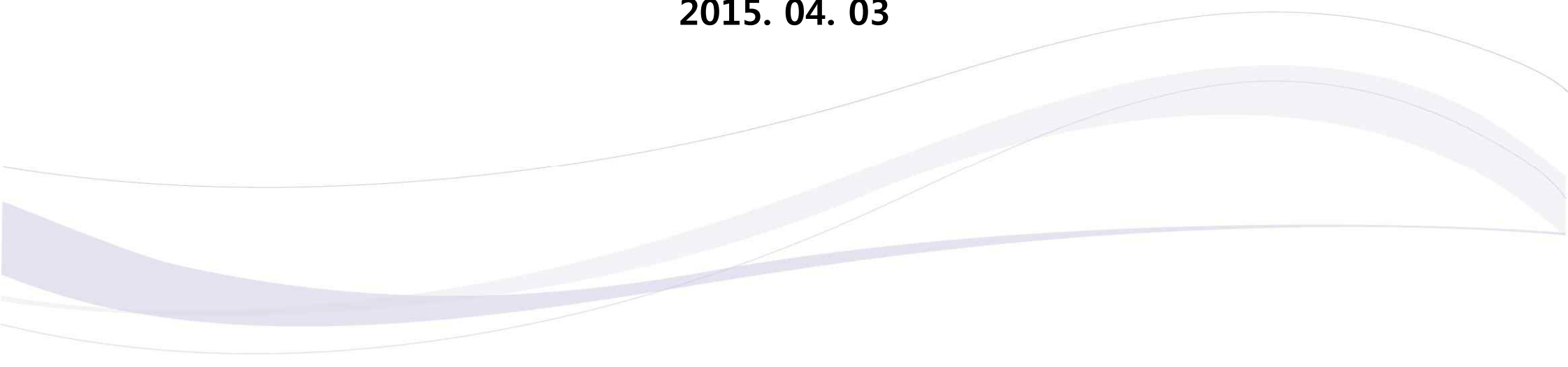


# 빅데이터 분석 사례와 가치 창출 방안

2015. 04. 03

The bottom of the slide features several decorative, overlapping wavy lines in shades of light blue and purple, creating a modern, abstract design.

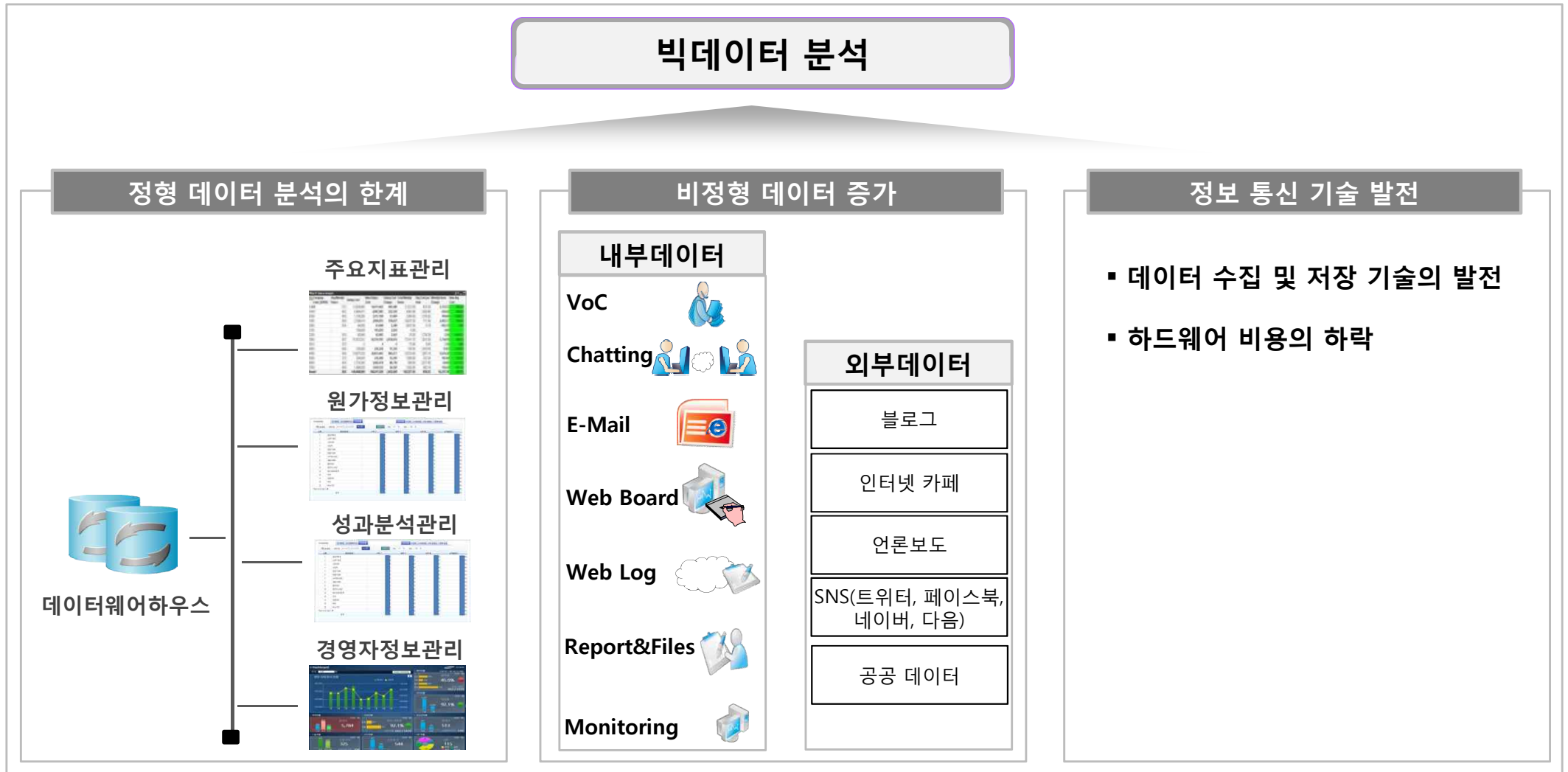
**I. 빅데이터 분석 개요**

**II. 빅데이터 분석 사례**

**III. 빅데이터 분석 업무**

**IV. BI 발전방향**

# 1. 빅데이터 분석 배경



## 2. 빅데이터 분석의 정의

업무 전문가의 주도하에 주위의 발생되고 있는 모든 데이터(비정형 포함) 분석 기반으로 선형적인 지식에 대한 검증 뿐만 아니라 의외의 사실로 미발생 데이터를 예측할 수 있는 분석

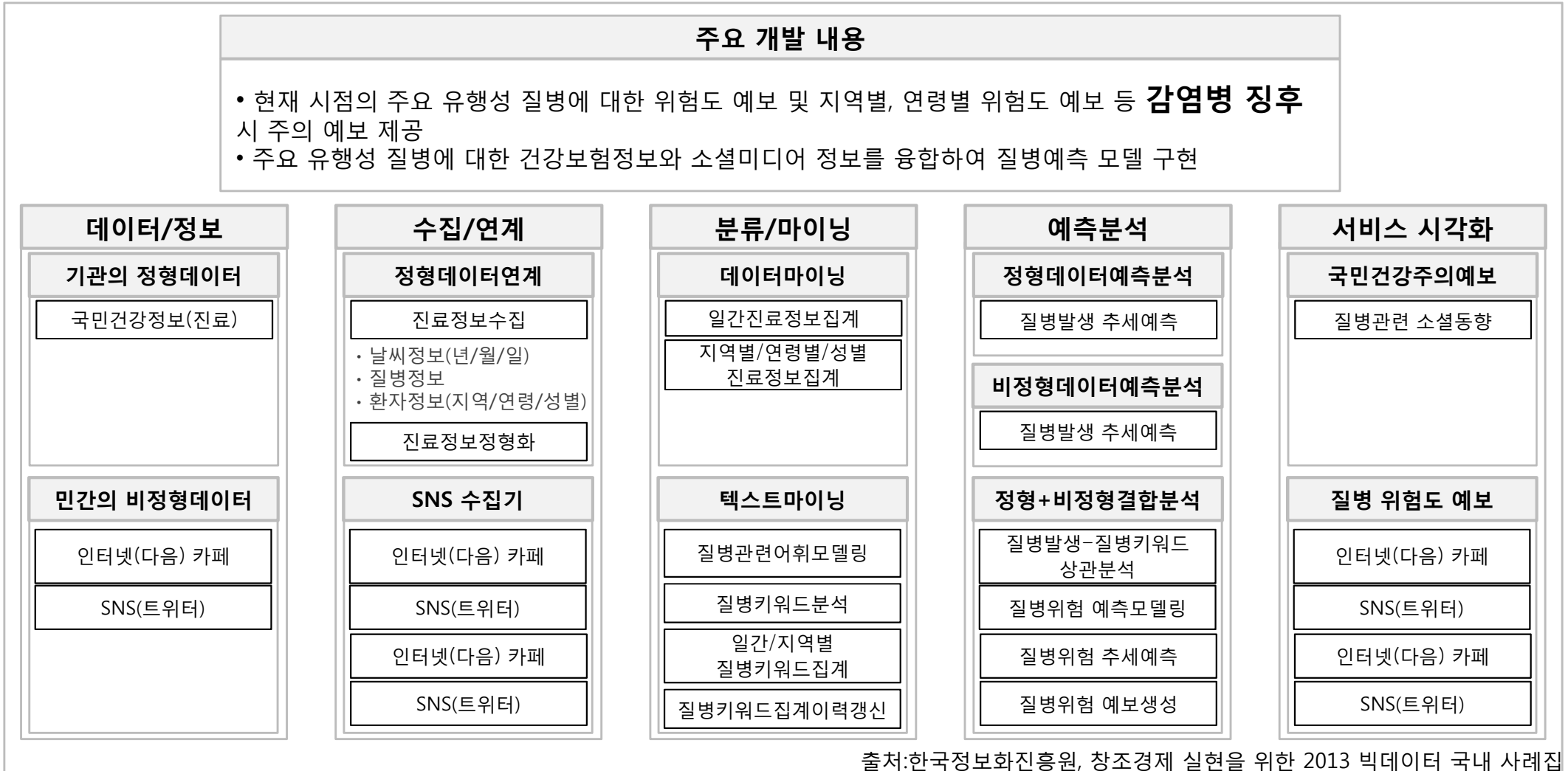


### 3. 발생 데이터 기반의 예측 분석

이미 추출된 발생 데이터를 기반으로 아직 발생 되지 않은 고객 활동에 대한 일정 구간의 유의 수준에서 예측하는 것

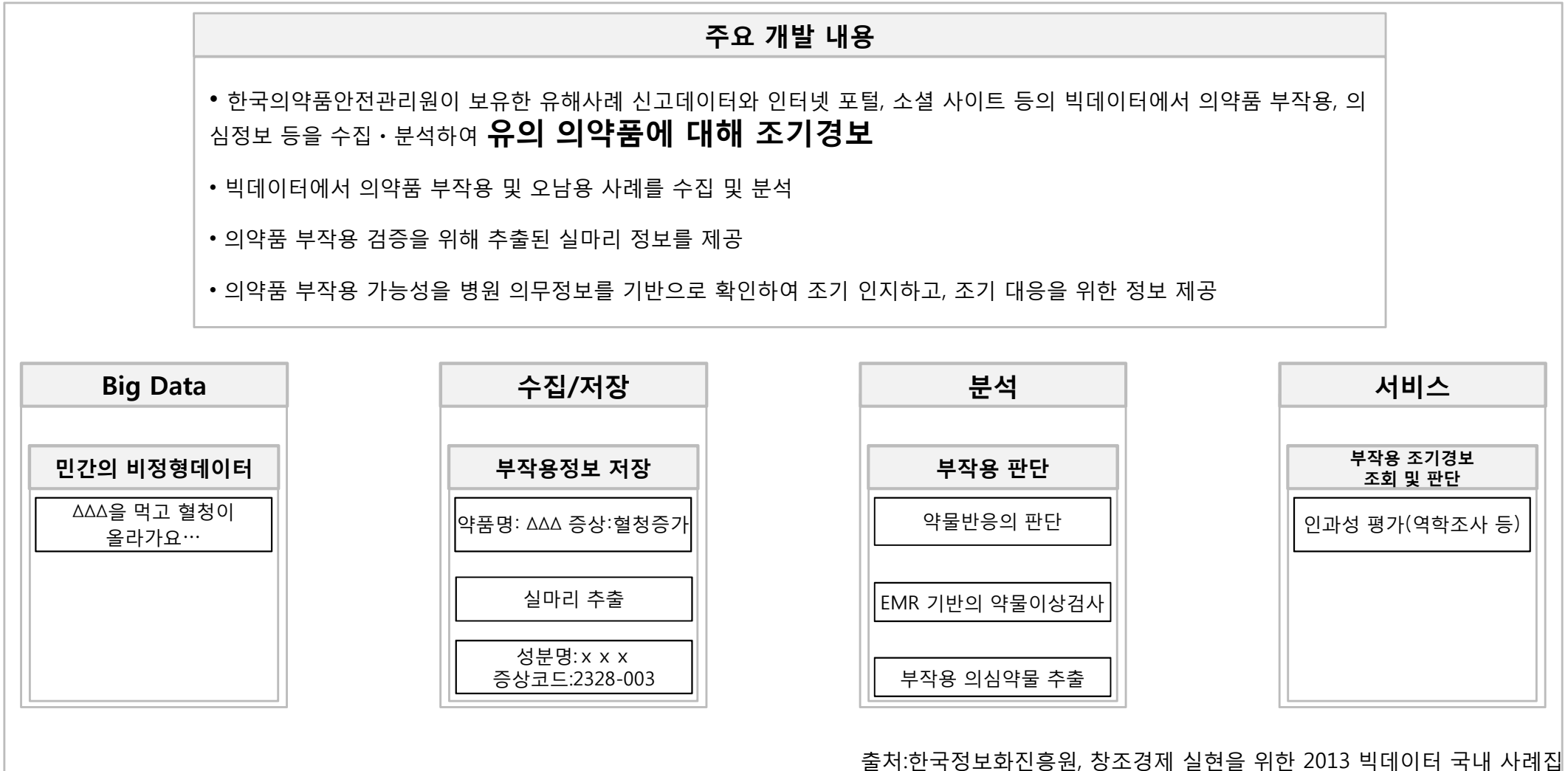


# 1. 국민건강 주의 예보 시범 서비스 구축



출처:한국정보화진흥원, 창조경제 실현을 위한 2013 빅데이터 국내 사례집

## 2. 의약품 안전성 조기 경보 서비스



### 3. 심실부정맥 예측, 입원병상 최적화 등 보건의료서비스

#### 주요 개발 내용

- 심실부정맥 예측 서비스 : 중환자로부터 수집되는 심실 모니터링 데이터를 빅데이터 플랫폼에 저장하고, 패턴 분석을 통해 **부정맥 발생 상황을 사전 예측**
- 입원병상 배정 최적화 서비스 : 병원이 보유하고 있는 환자 입원/진료 데이터, 입퇴원 기록, 보험 청구 데이터 등 병원 경영자료를 복합 분석하여 각 과별 병상 배정을 최적화 함
- 계절적 요인, 병원 의료진 수, 진료 행위의 변화를 모델에 반영하여 입원 병실의 가동률 변화 패턴을 파악

#### Big Data

##### 환자 모니터 기기

생체신호 데이터

##### 병원정형데이터

진료, 입퇴원기록,  
보험청구, 원무데이터

#### 수집/저장

##### 심실모니터링 데이터

생체심전도/심박수 등의  
생체신호데이터 저장

##### 병원정형데이터

진료, 입퇴원기록,  
보험청구, 원무데이터

#### 분석

##### 데이터 마이닝

심실부정맥예측

입원병상배정 최적화

#### 서비스

##### 사전예측

심실부정맥 경보/호출

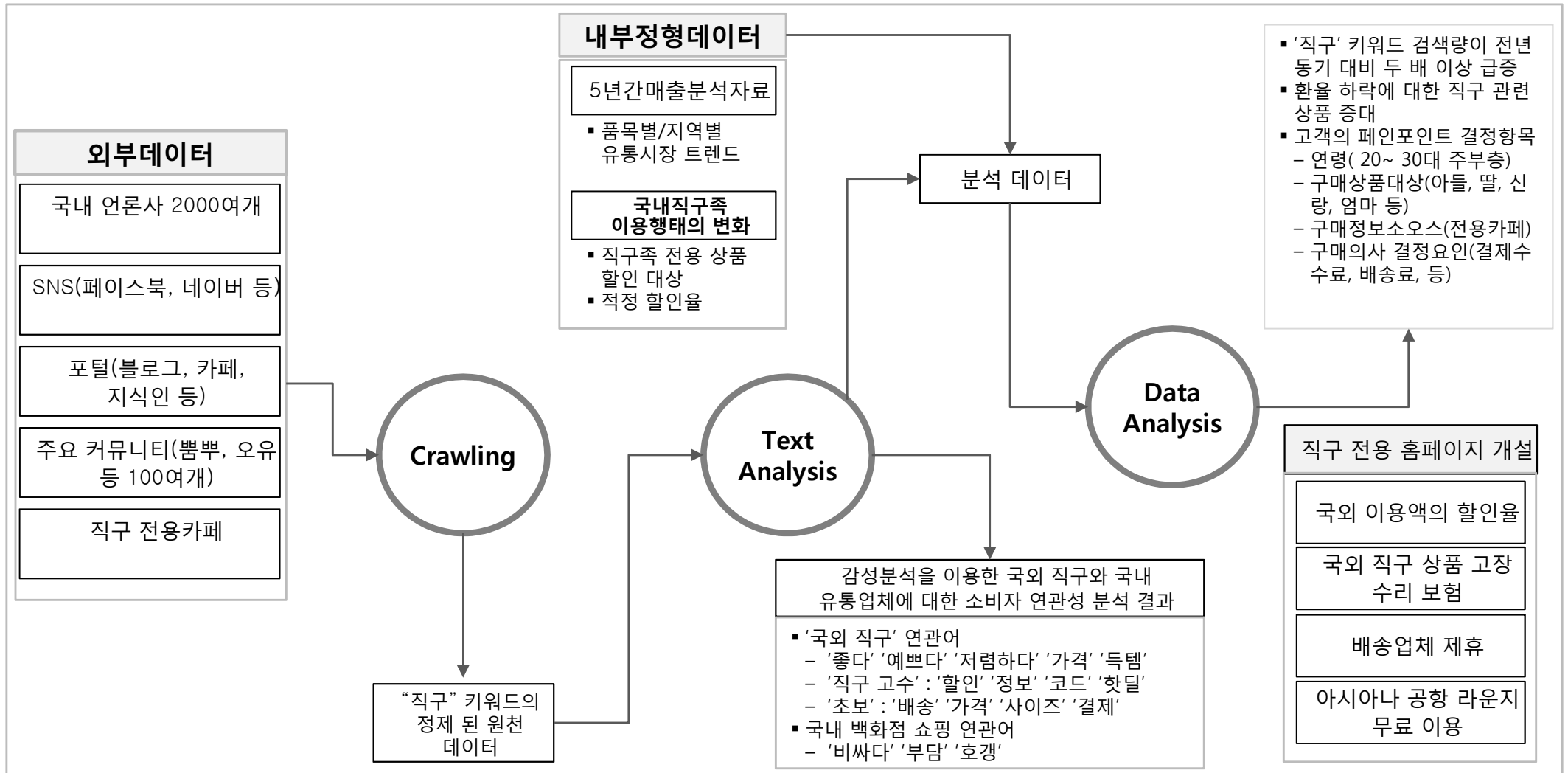
##### 최적화

입원병상배정최적화

출처:한국정보화진흥원, 창조경제 실현을 위한 2013 빅데이터 국내 사례집



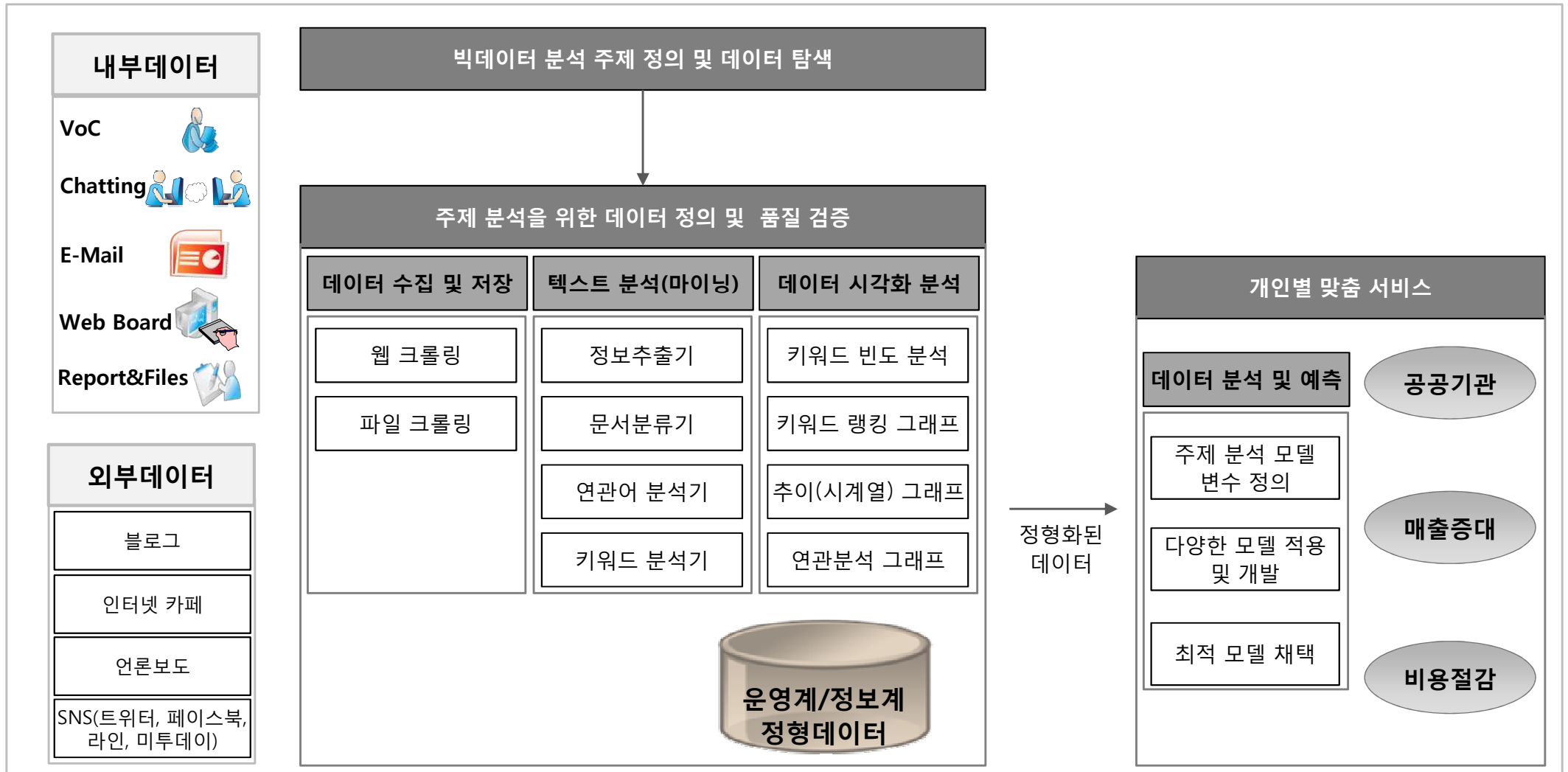
### 4. 신한카드 해외직구 전용카드 서비스



## 5. 빅데이터 분석 국내 사례 종합

빅데이터 활용사례	주관기관	활용데이터	적용기술	비고
국민건강 주의 예보 시범 서비스 구축	국민건강보험관리공단	<ul style="list-style-type: none"> <li>• 국민건강정보(진료내역)</li> <li>• SNS(트위트, 다음카페)</li> </ul>	<ul style="list-style-type: none"> <li>• 자연어처리기술을 이용한 텍스트 마이닝 기법</li> <li>• 통계분석/기계학습 기술을 이용한 데이터마이닝 기법</li> </ul>	
의약품 안전성 조기 경보 서비스	에스지에이 주식회사	<ul style="list-style-type: none"> <li>• 유해사례신고DB</li> <li>• SNS(블로그, 지식인, 카페, 트위터), 뉴스(웹)</li> </ul>	<ul style="list-style-type: none"> <li>• 자연어처리기술을 이용한 텍스트 마이닝 기법</li> <li>• 부작용 검증을 위해 코호트 기반 연구방법</li> <li>• 환자/대조군 비교 알고리즘 등</li> </ul>	
심실부정맥 예측, 입원병상 최적화 등 보건의료 서비스, 마약류 인지·감시서비스	서울아산병원	<ul style="list-style-type: none"> <li>• 환자심전도/심박수, 진료, 입퇴원기록, 병원 경영 데이터</li> <li>• SNS, 검색데이터</li> <li>• 국과수 마약류 관련 DB</li> </ul>		
신한카드 해외직구 전용카드 서비스	신한카드	<ul style="list-style-type: none"> <li>• 신한카드 내부 매출데이터</li> <li>• SNS(언론사, 포털 등)</li> </ul>	<ul style="list-style-type: none"> <li>• 자연어처리기술을 이용한 텍스트 마이닝 기법</li> </ul>	

# 1. 빅데이터 분석 개요도



## 2. 텍스트 마이닝 - 키워드 & 연관어 분석 예시

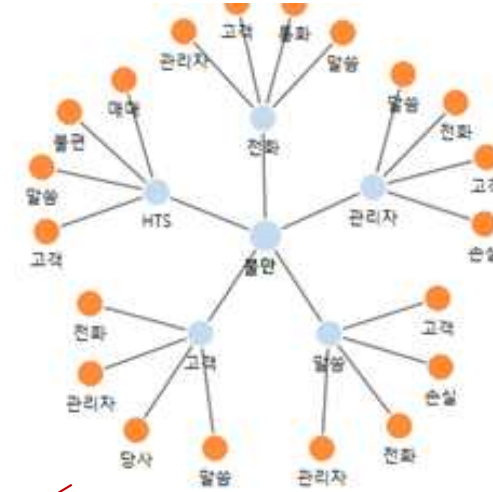
웹페이지에서 형태소 분석 기반의 키워드 분석

췌장질환	담즙산	이상중세	유해세균	대장암
이연경	성황리	방광암	검단탑병원	소화기질환
후두암	요로감염증	항암식품	폐암	뇌종양
결장	예고	배변습관	대변	위암
피부암	식도암	멋진	목소리	난소암
자궁암	발병율	체중감소	사마귀	직장암

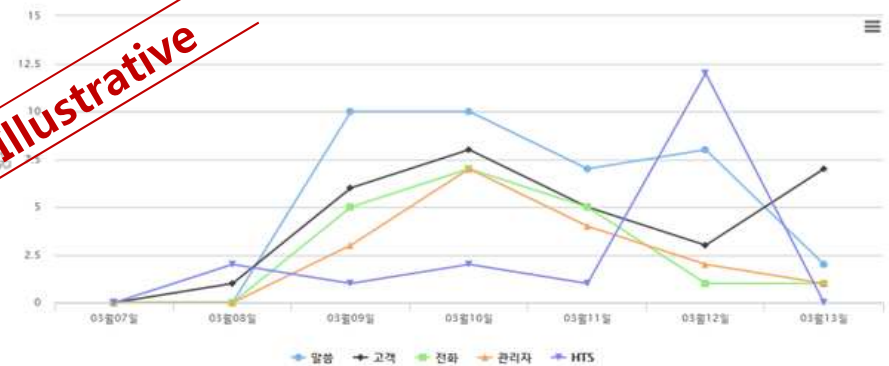
제목	요약	URL
대장암 말기 증상이 나타났다면? 'DOcale naver.com/cancerfight/254...	50세 이상의 성인남녀라면 5-10년마다 대장내시경검사 또는 대장이중조영검사를 하도록 권장하고 있습니다. 초기의 <b>대장암</b> 과는 달리 말기에는 배변습관의 변화, 혈변, 체중감소 등의 <b>대장암</b> 말기 증상을 호소하게 됩니다. 실제로 이와 같은 <b>대장암</b> 증상들이 나타나서 병원에 찾아갔지만 <b>대장암</b> 말기 증상으로 진단을...	http://cale.naver.com/cancerfight/2543
성황리에 마무리된 대장암의 날 행사, 검단탑 'DOcale naver.com/gumclantop/2388	방법, <b>대장암</b> 의 조기경전의 중요성, <b>대장암</b> 의 치료법을 직접 체험관을 통해 알기쉽게 설명하였습니다. 대학병원 교수들이 뽑은 '암수술분야 한국의 명'에 선정된 검단탑병원 이창환 과장! 올바른 식습관을 알려주는 부스도 마련되었습니다. 소화기암센터의 전문상담사가 무료건강상담을 실시하였습니다. 영양팀에서 직접...	http://cale.naver.com/gumclantop/2388
종합병원 건강 정보 'DOcale naver.com/dlttkfd/7/80	' <b>대장암</b> *대장염 *대장염 상담 *대장폴립 아토피상담실 서울 청철한의원 오리엔 사용전후 사진들 아토피상담실 서울 청철한의원 원라 *라이에중추근 *레이저박피술 *렘토스피라중 *라임병 *레이노병 *레이노 중후상담 *류마티스성관절염1 *류마티스성관절염2 *류마치스관절염상담 *류마티스열 *류마티스 *루푸스(홍반성창)...	http://cale.naver.com/dlttkfd/7/80

Illustrative

분석된 키워드 기반의 연관어 분석



Illustrative



## 2. 텍스트 마이닝 - 정보추출 예시 (1/1)

<육안소견 >

#1. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.4cm 이다. 전부 포매.  
 #2. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.7cm 이다. 전부 포매.  
 #3. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.6cm 이다. 전부 포매.  
 #4. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.4cm 이다. 전부 포매.  
 #5. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.4cm 이다. 전부 포매.  
 #6. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.7cm 이다. 전부 포매.  
 #7. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.3cm 이다. 전부 포매.  
 #8. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.4cm 이다. 전부 포매.  
 #9. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.3cm 이다. 전부 포매.  
 #10. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.7cm 이다. 전부 포매.  
 #11. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.7cm 이다. 전부 포매.  
 #12. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.3cm 이다. 전부 포매.  
 #13. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.5cm 이다. 전부 포매.  
 #14. 포르말린에 고정된 1개의 침 생검 조직으로 길이 1.5cm 이다. 전부 포매.

< MICRO (3 HE) >  
 DIAGNOSIS :  
 Prostate, #1-#14, needle biopsy :  
 1. ADENOCARCINOMA, Gleason's score 7 (4 + 3)/10 (#4)  
 (tumor length / core length ; 0.2/1.4cm)  
 2. ADENOCARCINOMA, Gleason's score 7 (3 +

Core No	(+)BxCore	%Positive Core	Max%Core	Tumor %core#1	core#1 Cancer 유무	core#1 GS1	core#1 GS2	core#1 Tumor lngth
14	7	50	76.92	0	무	0	0	0

Gleason Score ÷ No Of Core

Max(tumor lngth/core lngth#1,,,,,#14)

· 비정형 텍스트에서 분석 데이터 항목과 값들을 추출

Illustrative

## 2. 텍스트 마이닝 - 정보추출 예시 (2/2)

[검사 결과 자료 예시]

1-3. Stomach, subtotal gastrectomy:

Early gastric carcinoma

1. Location : lower third  
center at antrum and posterior wall

2. Gross type : EGC type IIc+IIa

3. Histologic type : tubular adenocarcinoma, moderately differentiated

4. Histologic type by Lauren : intestinal

5. Size : 1.2x0.5x0.1 cm

6. Depth of invasion : extension to submucosa (sm1) (pT1b)

7. Resection margin: free from carcinoma  
safety margin: distal 1.4 cm, proximal 12.5 cm

8. Lymph node metastasis : no metastasis in 46 regional lymph nodes (pN0)  
(0/46: lesser curvature, 0/6; greater curvature, 0/10; "1", 0/2; "4sb", 0/1;  
"5", 0/0; "6", 0/6; "8", 0/1; "8a", 0/1; "7", 0/2; "9", 0/7; "11", 0/2;  
"11p", 0/3; "12a", 0/5)

9. Lymphatic invasion : not identified

10. Venous invasion : not identified

11. Perineural invasion : not identified

비정형 텍스트의 구조화

[항목 정의서 예시]

항목	적용 여부	룰 Or 로직	ASIS화면번호
주진단명	Y	Stomach, subtotal gastrectomy: 다음의 텍스트	병리검사결과지
종양위치1	Y	" Location: " 다음의 텍스트 값	병리검사결과지
종양위치2	Y	" Location: " 다음 줄의 "Center at"이후의 텍스트 값	병리검사결과지
육안형	Y	"Gross type : " 이후의 값 (코드값은 임의적인 판단이 있음)	병리검사결과지
Histology	Y	" Histologic type: tubular adenocarcinoma, " 다음의 값	병리검사결과지
Lauren	Y	" Histologic type by Lauren : " 다음의 값	병리검사결과지
종양크기(max)	Y	"Size:" 다음의 "X"를 기준으로 분리하여 최대값	병리검사결과지
종양크기(min)	Y	"Size:" 다음의 "X"를 기준으로 분리하여 최대값	병리검사결과지
depth of invasion	Y	"Depth of invasion:" 다음의 값	병리검사결과지

[결과 예시]

주진단명	종양위치1	종양위치2	육안형	Histology	Lauren	종양크기 (Max)	종양크기 (Min)	Depth of Invasion
Early gastric carcinoma	lower third	antrum and posterior wall	EGC type IIc+IIa	moderately differentiated	intestinal	1.2	0.1	extension to submucosa (sm1) (pT1b)

### 3. 텍스트 마이닝 정보 추출 결과 시각화 예시



## 1. 빅데이터 분석 데이터 관리

## 데이터 클리닝 이슈

- 검사소견 판독 Text Data의 입력포맷 템플릿화를 통한 텍스트 정보 수집 결과 quality 증가 필요
- 검사결과 데이터 중 수치형 결과도 문자형으로 입력
- 수치형 결과 칼럼에 수치형/문자형 변수가 혼재
- 시점 변화에 따른 검사코드의 이력 관리

## 비정형 데이터의 중요성

## 일반정보

- 연령, 키, 몸무게, 복위 등

*Illustrative*

## Medical Cond.

- 흡연력, 흡연양, 금연기간, 고혈압여부 등

## 비정형 텍스트에서 추출된 데이터의 정보력이 높아짐

## Trus-Bx 정보

- 전체 Core 수, 양성 Core 수, Core Tumor 길이 등

## PreOP MRI 정보

- Size, Shape 등

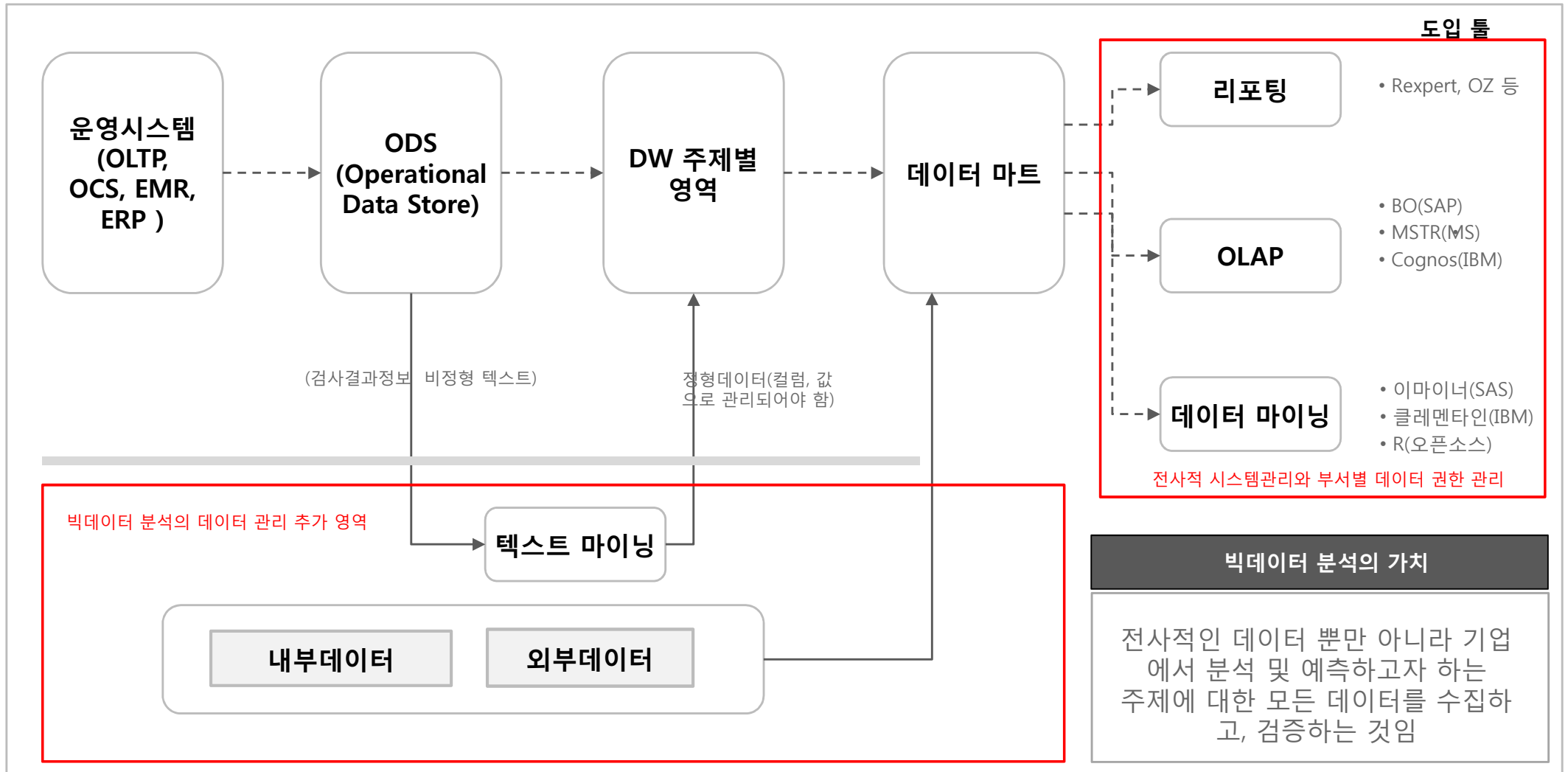
## Pathology 정보

- Size, Stage 등

*Illustrative*



## 2. BI에서 빅데이터 분석으로의 발전



***End Of Document***