

OWLNEST® SE-Mainframe

Proactive discovery and insights from contexts

OWLNEST Corp.

주소 : 서울특별시 강남구 자곡로 174-10 강남에이스타워 G9 318호 (06373)

문의 : contact@owl-nest.com, +82 2-742-3021

OWLNEST® SE-Mainframe System Requirements

To learn more about OWLNEST® SE-Mainframe, please visit <http://owl-nest.com/technology>, test the applications and see the references.



기능항목 조건표 (1/2)
 (기본) 검색 엔진 영역
 (기본) 형태소분석 엔진 영역
 (선택) 텍스트분석 영역

구분	항목	항목 상세
텍스트 분석	검색 엔진 영역	키워드 조건 검색
		- 검색 기간 설정 기능
		- AND, OR, NOT 복합 검색 기능
		- 1 차 검색 결과 기반 2 차 검색 기능
		- 검색 결과의 기간별 빈도, 트렌드 시각화 (일/주/월) 기능
		2개 이상의 선택 키워드 검색 결과 비교
		- 비교 대상 키워드의 검색 결과 조회 기능
		검색 결과 텍스트 원문 조회 기능
		- 검색 결과 텍스트 원문 조회 기능
		불용어 처리
		- 불용어 등록을 통해 무의미한 단어 검색 제외 기능
		사용자 검색 단어 기간별 통계
		- 주요 사용자 검색 단어 리스트 조회 기능
	신출 키워드 추출	
	- 신출 단어 중 일정 빈도/증감 이상의 단어 추출 기능	
	- 기간별 비율증가 단어 추출 기능	
	- 기간별 절대 빈도 상위 단어 추출 기능	
	형태소분석 엔진 영역	사전 모듈화
		- 제품 관련 단어 형태소 사전 등록 기능
		- 단어별 분석 기준 개체명 분류 등록 기능
언어별 엔진 모듈화		
- 언어별 형태소분석 엔진 추가 기능		
자동 띄어쓰기 모듈		
- 한글 텍스트 자동 띄어쓰기 필요 시 적용		

기능항목 조건표 (2/2)

(기본) 검색 엔진 영역

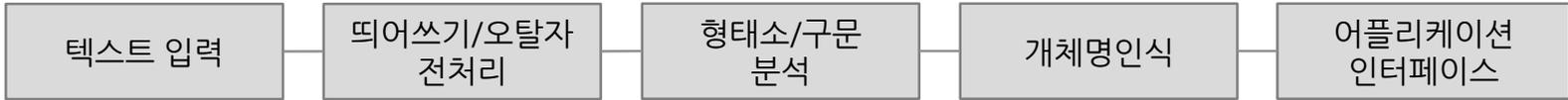
(기본) 형태소분석 엔진 영역

(선택) 텍스트분석 영역

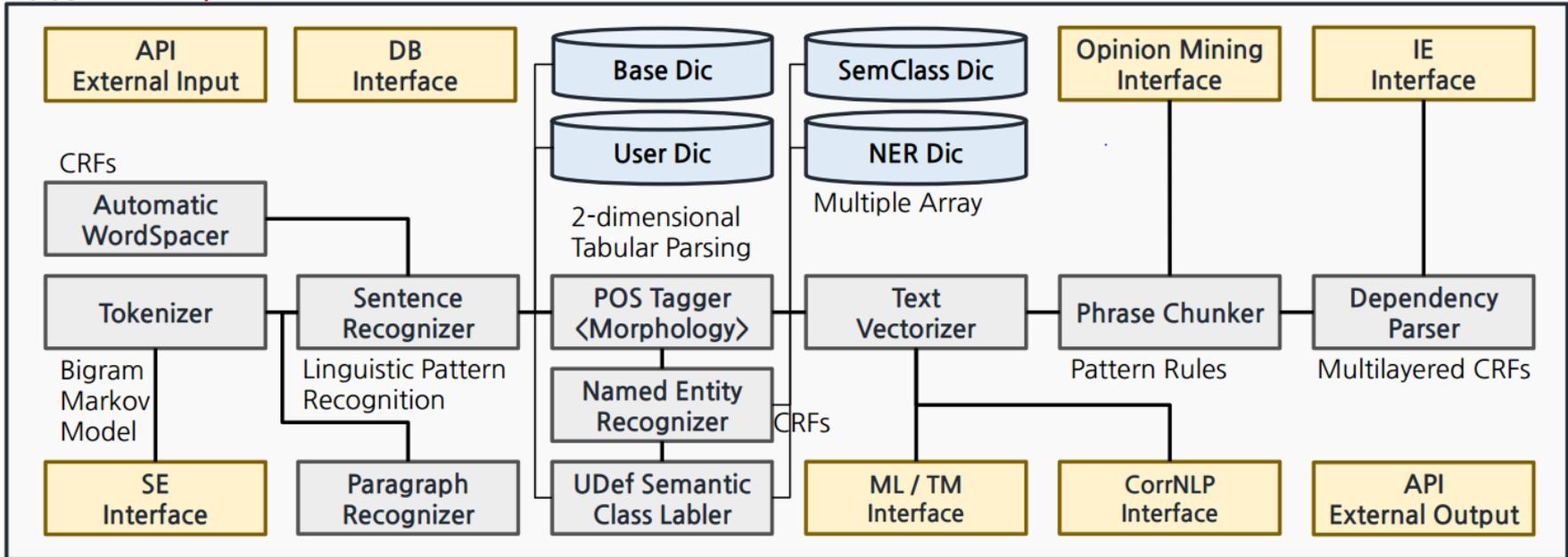
구분	항목	항목 상세
텍스트 분석	텍스트 분석 영역	키워드 기반 버즈량 분석
		- 기본 감성 (호감, 비호감, 의문) 기준 버즈량 조회 및 트렌드 시각화 기능
		- 감성 품질 주요 요소 기준 버즈량 조회 및 트렌드 시각화 기능 (타사비교)
		- 데이터 소스별 점유율 계산 결과 조회 및 시각화 기능
		- 버즈, 판매량, 기본 감성 간 상관관계 조회 및 시각화 기능
		검색 키워드의 연관 단어 간 Word Cloud, 키워드 네트워크 분석
		- 워드 클라우드, 키워드 네트워크 시각화 기능
		- 연관 통계 모델 기준 컨셉 키워드 그래프 시각화 기능
		검색 결과 텍스트의 키워드 마이닝 결과 제공
		- 검색 키워드와 관련된 주요 키워드 클러스터 조회 및 시각화 기능
		오피니언 리더 분석 기능
		- 오피니언 리더 중심 기본 감성, 감성 품질 언급 휴먼 네트워크 분석, 시각화
		감성 품질 주요요소 기준 정보 추출
		- 감성 품질 주요요소 관련 감성 키워드 추출
- 감성 품질 주요요소 관련 세부항목 추출		

SE-Mainframe은 텍스트 데이터 전처리를 위해 NLP-Mainframe 을 기반으로 견고한 텍스트 분석 통합 아키텍처를 제공합니다.

자동 띄어쓰기, 오타자교정, 문장/문단경계 인식, 형태소분석, 구문분석, 개체명인식
 인터페이스 : 검색, 연관어, 기계학습, 감성분석



처리량 : 1 MB/sec, 1-Core



SE-Mainframe 의 주요 구성요소와 처리 흐름은 아래와 같습니다.

- 분석 프레임워크 마련을 위한 검색 기반 아키텍처를 제공합니다.
- 형태소 분석 모듈, 자동 띄어쓰기 모듈, 디스크/인메모리 인덱싱 모듈 지원



Admin Center

- 문서 입력
- 검색필드 구분
- 기간필드 구분
- 05 검색 인터페이스

SE-Mainframe

01 토크나이저 (언어별)

- Tokenizing Engine
- 일반/복합 키워드 처리

02 인덱싱

- Statistics Analyzer
- Inverted Indexing
- In-Memory Indexing

07

- Index Seg 1
- Index Seg 2
- Index Seg 3
- Index Seg 4
- Index Seg 5
- ...

04 랭킹/캐싱

- 검색결과 랭킹
- 검색패턴 학습

03 인덱스 관리

- 인덱스 생성
- 인덱스 저장
- 인덱스 백업

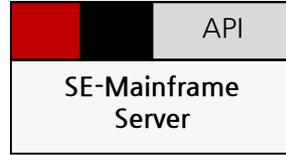
06 검색 결과

- 검색결과 조회
- 검색옵션 설정

- 고성능 한글 Tokenizer와 병렬색인 자료구조를 통하여 관련된 원문에 고속 도달
- 초고속 토크나이저, Tensor Memory 구조활용 연산 및 연관 검색결과 정확도 랭킹
- 인덱스 대상 파일 생성/저장 및 안정성을 위해 주기적 백업
- 검색결과 문서 랭킹, 결과 캐싱을 통한 검색속도 안정화
- 검색대상에 해당하는 문서DB 입력 인터페이스
- 관리자 UI를 통한 시스템의 검색결과 조회 및 검색옵션 설정
- Index Seg 저장소 파일 관리 및 In-Memory Indexing

SE-Mainframe 의 주요 특징은 아래와 같습니다.

- 시스템을 위한 서버 설치형 소프트웨어
- 서버의 하드웨어 성능을 극대화할 수 있는 기업형 솔루션



- 개발자를 위한 클라이언트 API
- 각 컴포넌트 모듈에 대한 외부 제어 가능형 사용자 솔루션

다중형식 처리

- 다양한 형식의 작업 문서, 입력파일 자동변환 기능

통합 인터페이스

- 컴포넌트 모듈별 실행 및 상태 조회를 위한 사용자 인터페이스

메타정보 분석

- 단어/구문 수준의 정보를 자동 분석하여 메타정보화

유니코드 인코딩

- 다국어 및 기호 문자열의 처리를 위한 최적의 인코딩

웹스케일 처리

- 대용량 파일의 처리를 위한 멀티 프로세싱 기능 지원

취급 파일 (pdf, doc, ppt, hwp, xls, txt) 의 통합 저장, 메타정보 관리

csv, json, xml 등의 파일출력 및 DB 인터페이스 지원



통합 자료 구조

컴포넌트 통합형 입출력 제어구조로 자료 관리의 편리성

사용자에게 익숙한 엑셀 테이블 형태의 자료 관리 구조



솔루션 제어 기능

중요 키워드에 대한 일반/특수/확장 개체명 정보 자동 인식

개체명-표현 간의 의존관계의 자동 추출을 통한 메타정보화



구조적 정보

유니코드 처리를 통한 텍스트 데이터 저장/관리

언어별 텍스트 인코딩 통합 관리/저장/추출



최적 인코딩

웹스케일의 입력 텍스트 처리를 위한 고속 병렬 문자열 토큰화 기능

토큰별 인덱스 생성과 메타정보의 연결을 통한 빠른 정보처리



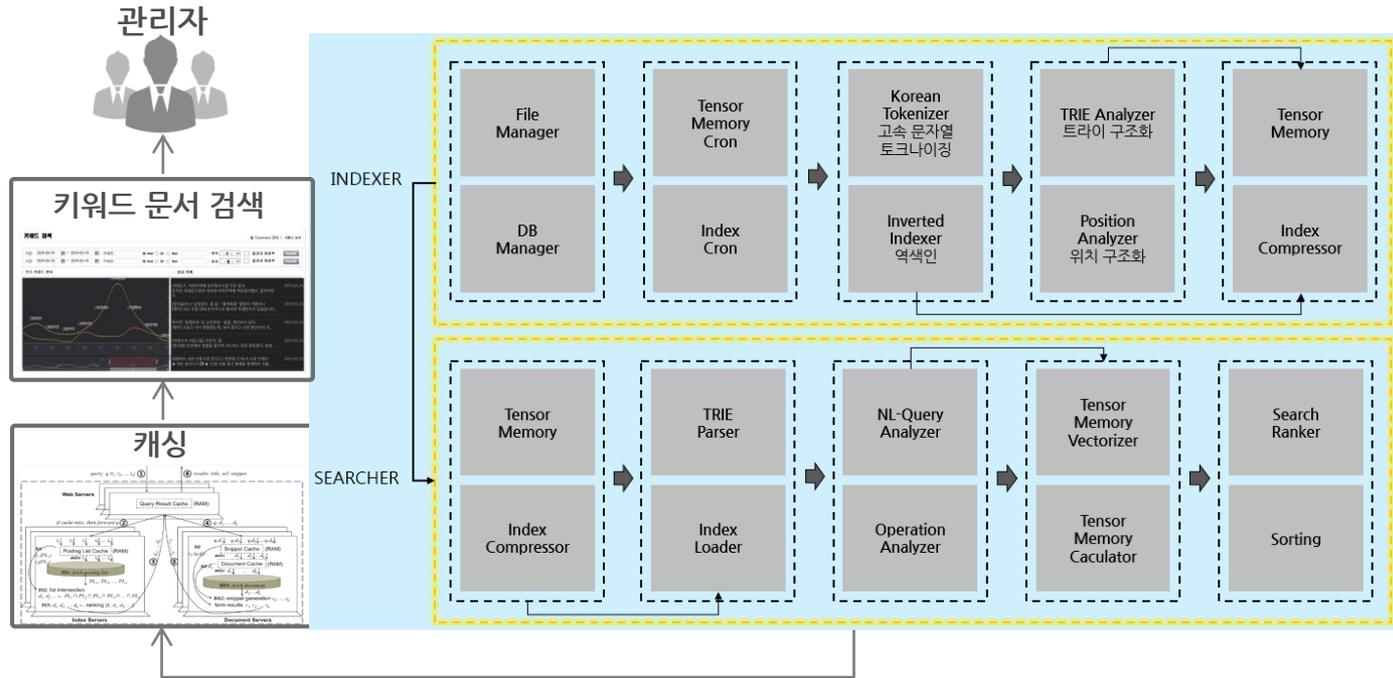
고속 병렬 처리

검색엔진, 키워드 검색을 통해 결과 텍스트 원문을 제공합니다.

- 키워드 검색 기능은 색인엔진을 이용하여 대규모 데이터로부터 빠른 검색결과 제공
- 고성능 Tokenizer와 병렬색인 자료구조를 통하여 사용자가 찾는 원문에 고속으로 도달
- 복합입력 검색 기능 지원, 자연어 검색과 같은 사용자 입력 인터페이스의 편리성을 제공
- 검색 결과의 캐싱을 지원, 순간검색빈도와 항시검색빈도의 반영으로 안정적 속도 지원

텍스트 원문 검색 기능

- 키워드 입력 검색
 - 일반 키워드 입력 및 검색 기능
- 복합입력 검색
 - 복합명사 인덱싱 : 색인어 개수 감소
 - 명칭 색인어 추출 : 개체명
 - 입력질의 형태 무관 처리 기능
- 색인 엔진
 - 유니코드 색인파일 생성
 - 디스크/인메모리 자료 관리
 - 로딩/인덱싱 속도 고속화
 - 전역메모리 최적화
 - 초고속 토큰라이저 (1MB/sec)
 - Tensor Memory 구조를 이용한 근접연산 및 연관도 검색결과 정확도 랭킹
- 캐싱 지원
 - 순간검색빈도와 항시검색빈도의 측면에서 검색속도 안정화
 - 검색에 관련된 각 데이터 플로우에 대한 캐싱을 통해 속도 및 자원 최적화

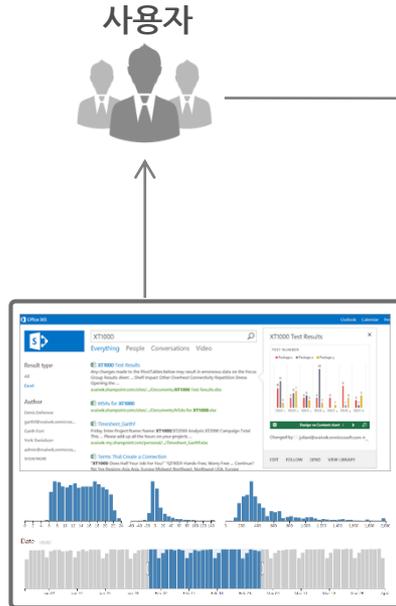


검색엔진, 다양한 조건의 키워드 검색 기능을 제공합니다.

- 키워드 복합 검색 (AND, OR, NOT) 논리연산자 처리 기능 제공
- 1차 검색 결과 기반 2차 (결과내) 재검색 기능을 지원
- 기간설정 검색을 위해서, 검색 기간 설정 및 날짜 기준 정렬 기능 지원
- 검색결과와 기간별 (일/주/월) 빈도표시를 위한 통계데이터 제공

다양한 조건의 복합 검색 기능

- 키워드 복합 검색
 - 자유로운 질의 키워드 입력 검색
 - Set operation 기반 논리연산
 - 색인엔진을 이용한 복합키워드 고속검색
- 결과내 재검색 기능
 - 1차 검색결과 기반 결과내 재검색
 - 순간검색빈도 재검색 통계를 이용한 2차 캐싱 구조를 통한 자원 최적화
- 기간설정 검색
 - 검색기간 설정 조회 기능
 - 최근 문서 순서대로 소팅 기능
 - 최근시점 기준 기본 설정
- 기간별 (일/주/월) 빈도표시
 - 기간별 빈도 표시를 위한 주요 색인 키워드 고빈도순 JSON 형태 데이터 제공
 - 기간별 중요도 표시를 위한 주요 색인 키워드 중요도순 JSON 형태 데이터 제공



사용자



Four search interface examples are shown, each with a checkmark and a title:

- ✓ 키워드 복합 검색: Search interface with 'And Or Not' radio buttons and a '결과내 재검색' checkbox.
- ✓ 결과내 재검색: Search interface with a highlighted '결과내 재검색' checkbox.
- ✓ 기간설정 검색: Search interface with date range selection and a '결과내 재검색' checkbox.
- ✓ 기간별 (일/주/월) 빈도표시: Search interface with a date range selection and a '결과내 재검색' checkbox.

검색엔진, 다중 키워드 검색결과 비교 기능을 제공합니다.

- 일정 기간에 대한 다중(2개 이상) 키워드 동시조회 결과 제공
- 각 키워드에 대한 동시 빈도조회를 위한 라인차트 데이터 추출 제공
- 다중 키워드 포함 원문 리스트 확인 및 해당 원문 조회 기능 제공
- 분석 결과별 시각화 인터페이스를 통해 편리한 UI 프레임워크 연계성 제공

다양한 조건의 키워드 검색

- 다중 키워드 동시조회
 - 자유로운 질의 키워드 다중입력 검색
 - Set operation 기반 논리연산 병렬처리
 - 색인엔진을 이용한 다중키워드 고속검색

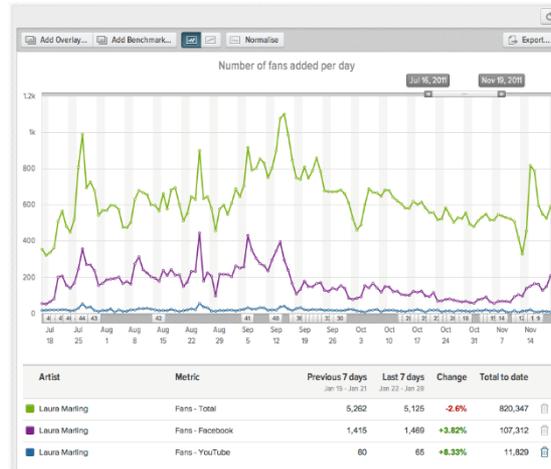
- 동시 빈도조회 라인차트 데이터
 - 다중 키워드 동시조회를 통한 키워드 별 동시 빈도조회 라인차트 데이터 제공

- 해당 원문 조회 기능
 - 검색기간 설정 조회 기능
 - 최근 문서 순서대로 소팅 기능
 - 최근시점 기준 기본 설정

- 분석 결과별 시각화 인터페이스 제공
 - 각 분석 결과별 UI 인터페이스 지원
 - 시각화를 위한 JSON 형식 데이터 제공



✓ 다중 키워드 동시 조회



✓ 키워드 검색 관련 해당 원문 확인 조회

Top links

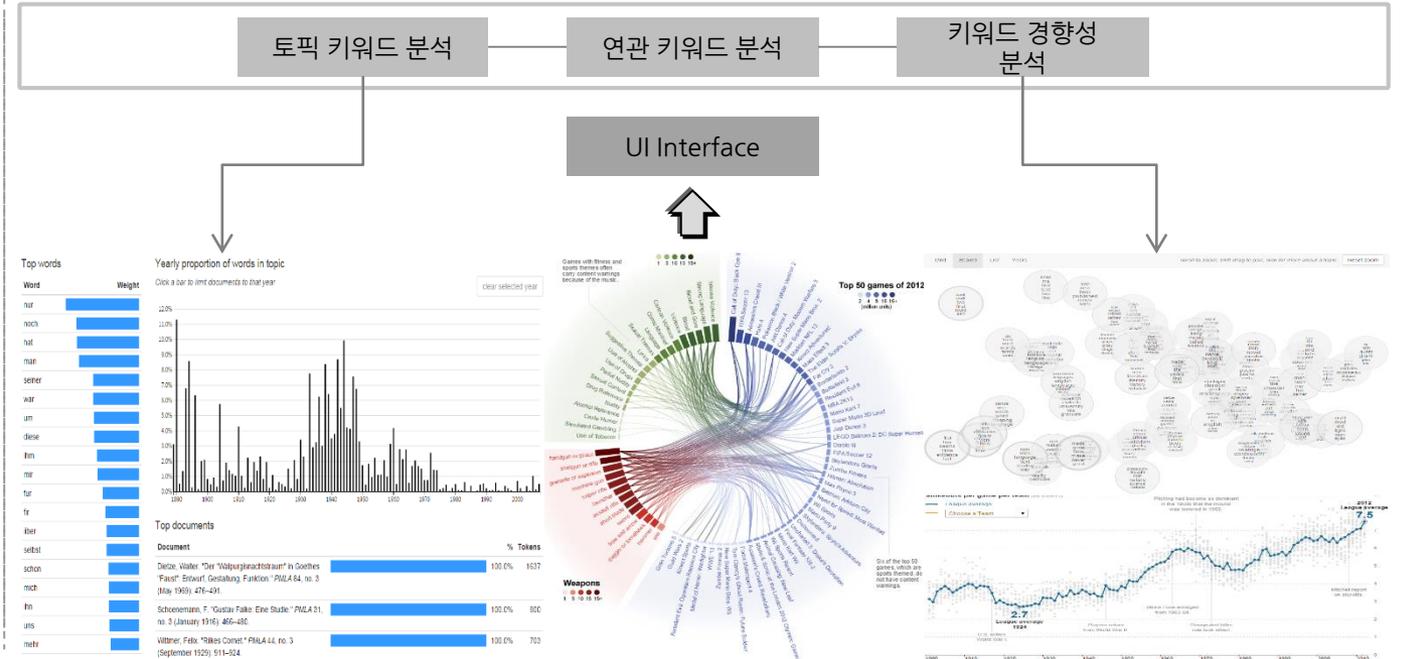
- Ranking the decision makers: Who is best at the NFL draft ...
- SS who really stands the best chance to nab the most from its rookies now that the 2010 NFL draft has concluded?
- Revisiting the Ferentz-Thesset COY debate - Big Ten Blog - ESPN
- Last week's NFL draft rekindled a hot topic on this blog - the 2009 Big Ten Coach of the Year race between Iowa's Kirk Ferentz and Ohio State's Jim T. ...
- Revisiting the Ferentz-...
- Ranked: Raiders - Al Davis His Best NFL Draft in Years ...
- Finally, Al Davis and his entourage pulled off a solid, respectable outing in New York during last weekend's NFL draft and it's almost ...
- Revisiting the Ferentz-...
- NFL Ticket Exchange: Roger Goodell: NFL draft will keep three-day format ...
- Facebook • 9 minutes ago
- NFLivExchange: Roger Goodell: NFL draft will keep three-day format ow.ly/1FLJ
- Twitter • 9 minutes ago
- City Rights: Zeno Abrahams - Caleb Campbell, 2009 NFL Draft ...
- Visit YouTube | Visit UShow.com Viewers of this biggest videos from the NFL Draft will remember the All-American ht Caleb Campbell, ...
- http://www.sfgate.com/cgi-bin/.../rdad17... All mentions »
- Scout.com: 2010 NFL Draft Superlatives Scout.com's NFL Draft Analyst Chris Steuber identifies the best, the worst and the sleeper selections from last week's NFL Draft http://profootball.scout.com/290798.html All mentions »
- stadiumspod: Worth the wait: The new NFL Draft format, going from two days to three days, meant just one thing to rookie running ... bit.ly/0b7109

검색엔진, 키워드 트렌드 분석을 제공합니다.

- 문서 내 트렌드 분석을 위해 주요 키워드 및 가중치 키워드 제공
- 각 주요 키워드에 대한 문서내 2차 키워드 및 주요 키워드 간 연관관계 통계적 자동 추출
- 키워드 경향성 분석 기능을 통해 시기별 키워드의 변화 양상을 한눈에 확인
- 분석 결과별 시각화 인터페이스를 통해 편리한 UI 프레임워크 연계성 제공

키워드 트렌드 분석

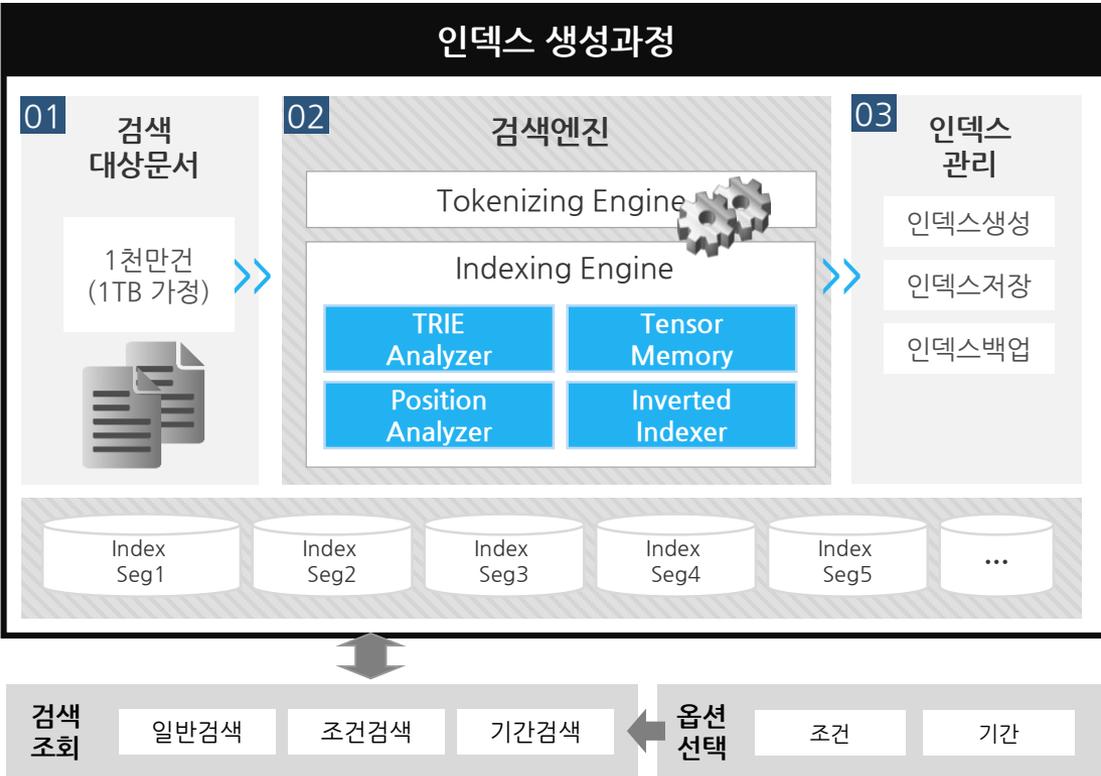
- **트렌드 키워드 분석 기능**
 - 문서 내 주요 토픽 분석을 통한 주요 키워드 및 관련문서 추출
 - 중요 키워드 추출 알고리즘을 이용한 가중치 키워드 목록 추출 제공
- **연관 키워드 분석 기능**
 - 토픽 키워드 간 문서내 2차 키워드 연관도 (Correlation) 통계 측정
 - 연관도 추출 알고리즘을 통한 토픽 키워드 간 연관관계 트렌드 분석
- **키워드 경향성 분석 기능**
 - 토픽 키워드 별 기간별 경향성 조회
 - 기간별 절대빈도 상위 단어 조회
 - 기간별 비율증가 상위 단어 조회
- **분석 결과별 시각화 인터페이스 제공**
 - 각 분석 결과별 UI 인터페이스 지원
 - 시각화를 위한 JSON 형식 데이터 제공



검색엔진 구동을 위한 HW 용량 산정은 아래와 같습니다.

- 검색엔진 대상파일 및 인덱스 파일을 관리하기 위해 적절한 HDD 공간 산정
- 인덱스 파일의 저장을 위해 입력 문서의 사이즈 대비 약 2.5배의 공간 요구
- 인메모리 인덱싱을 위해 고용량 램 환경 필요

인덱스 생성시, 원문 대비 용량 증가



주요 특징 / 내용

- 1 검색 대상문서 입력**
 - 고객 VOC 문서를 검색 대상으로 가정
 - (가정) Raw Data : 연간, 1천만건, 1TB
 - 검색엔진을 통해 인덱스 파일 생성시, Raw Data 기준 2.5배 정도의 사이즈를 요구
- 2 검색엔진**
 - 초고속 토크나이징을 통해 검색 인덱스 문자열 토크 생성
 - 인덱싱 엔진을 통해 인덱스 관련정보 형성
- 3 인덱스 관리**
 - 검색엔진을 통해 생성된 인덱스 파일 관리
 - (가정) Index 파일 : Raw Data 기준 1천만건 검색인덱스 처리시 2.5TB 인덱스 파일 생성
 - 안전한 관리를 위해 인덱스파일 백업 저장 옵션
- 4 운용상황으로 가정시 최소 6TB 추천**
 - Raw Data : 1 TB
 - Index File : 2.5 TB
 - Index Backup : 2.5 TB

SE-Mainframe 은 기능유형으로 구분된 각 사전유형을 모듈구조로 관리합니다.

● 모듈형 사전구조의 효과



도입 시스템의 안정화 주기 단축
빠른 시스템 안정화

손쉬운 사전관리 및 유지보수
유지보수 효율성 향상

요구사항에 따라 사전 재조합
사전생성 효율화

일원화 구조 모듈의 위험요소
오류등록 Risk 최소화

기능유형	사전유형								사전갯수
내용어	일반명사	고유명사	대명사	신조어	한자어	복합명사	개체명	차용어	12
기능어	부사	조사	어미	어근	접미사	접두사	수사	감탄사	20
용언	동사	형용사	보조용언	관형사	지정사				6
기타	부호	외국어	숫자	한자	사용자정의				5
총계									41

SE-Mainframe 의 도입으로 향상된 검색결과 품질을 기대할수 있습니다.

- 문장간 문맥적 유사도 계산, 특정 키워드 중심 재검색, 입력 검색어 길이 무제한 지원
- 연관어/개체명/독음/동형이음 검색

일반 검색엔진

사용자 요구사항	입력 키워드를 포함하는 문서를 조회
입력 쿼리	키워드, 조건 (AND, OR)
검색결과 특징	입력 키워드가 포함된 문서 화면노출
랭킹 조건	날짜, PageRank
특징	컨텍스트 유사문서 검색 불가

NAVER 검색엔진

일반	원래 모든지 잘먹던 아빠인데 당뇨가 생긴뒤로는 밥을 포함한 모든 음식을 거의 안먹어요 그래서 지금 엄청 말랐어요ㅠㅠ 그리고 요즘 주변분들이 당뇨로인해서 합병증도오고, 병원에 입원하는 경우를 많이봐서 많이 걱정이되요 원래 당뇨가 있으면 입맛이 없는건가요?
1	그래서 사람의 몸 속에서도 수분 및 영양분을 조절하여... 떨어 뜨려주어서 당뇨에도 상당히 좋다. (참고 - S.B.S 잘.먹고.잘.사는.법 : 칩은 만능의 약재이며 음식이다)... 치매를 포함한 모든 형태의 치매 위험이 평균 25% 낮은 것으로 ...
2	그리고 오리편은 피지를 깨끗이 제거해줌으로, 몸 전체... 호두는 그냥 먹어도 맛있지만 호두에는 체내에 잘 흡수되지 않는 성분이 포함되어 있으므로 기름으로 짜서 먹거나... # 모든 병의 근본적인 치료는 음식에 있습니다 ...
3	그리고 콩류 음식 많이 드시면 좋습니다. 콩 중에서도... 호두는 그냥 먹어도 맛있지만 호두에는 체내에 잘 흡수되지 않는 성분이 포함되어 있으므로 기름으로 짜서... 죽염= 모든 생물이 썩지 않는 것은 염성의 힘 때문인데 ...

컨텍스트 검색엔진

사용자 요구사항	입력 문장의 컨텍스트 유사문서 조회
입력 쿼리	일반 자연어 문장/문서
검색결과 특징	입력 문장/문서와 유사한 문서 화면노출
랭킹 조건	날짜, 컨텍스트 유사도
특징	키워드 불포함 문서도 추출

컨텍스트 검색엔진

일반	원래 모든지 잘먹던 아빠인데 당뇨가 생긴뒤로는 밥을 포함한 모든 음식을 거의 안먹어요 그래서 지금 엄청 말랐어요ㅠㅠ 그리고 요즘 주변분들이 당뇨로인해서 합병증도오고, 병원에 입원하는 경우를 많이봐서 많이 걱정이되요 원래 당뇨가 있으면 입맛이 없는건가요?
1	당뇨로입원중이신아버지가있는데병원밥이입맛에맞지않는지잘 먹지도못하고그래서할머니께서김치랑젓갈무쳐놨다고가져다주라는데 아무리밥을못먹어도당뇨면김치면몰라도젓갈은짠음식이니까먹으면안되죠? ...
2	당뇨로 인해 급격한 체중변화가 있은후부터 이런 증상이 있었던거 같은데 다리근육이 뭉치는 느낌이 들면 양다리를 맘대로 못 가누게 되어 일어서지 못하고 꼼짝을 못하는 경험을 하게 되는데(이런 증상이 나타날땐 어김없이 당이 많이 함유된 ...
3	고혈압은 원래 있었고, 병원에 가서 당뇨 수치를 봤더니 360~370이 나왔네요 (엄청높다능) 그래서 걱정이 되서요 저는 아빠 없으면 못살아요 (죽는병은 아니지만) 그래서 아빠가 식단을 채소 위주로 바꾸고,운동을 하기로 했어요 ...

2015년 이후, 여러 고객사에서 비즈니스 니즈와 프로젝트의 성공적인 목표 달성을 위해 SE-Mainframe 을 도입했습니다.



프로젝트 명	발주처	기간	역할	형태
재난안전 맞춤형 텍스트 탐색 및 실시간 분석 UI 개발	국립재난안전연구원	2016. 6. ~ 2016. 12.	연구개발, 시스템개발	시스템 개발
삼성페이 서비스 고도화를 위한 분석 시스템	삼성전자	2016. 3. ~ 2016. 6.	분석 컨설팅, 솔루션 납품	분석 컨설팅
삼성그룹 컨설팅 영업지원 종합솔루션	삼성생명	2015. 6. ~ 2015. 12.	머신러닝 솔루션 개발	컨소시엄 개발
삼성생명 BDA 경영자원화 구축	삼성생명	2015. 6. ~ 2015. 12.	솔루션 납품 모듈 커스터마이징	컨소시엄 개발
텍스트마이닝을 이용한 교통정보체계 감사토픽 분석	감사원	2015. 2. ~ 2015. 4.	공동연구/개발	Term License
미래위험 변화예측을 위한 사회환경탐색기술 개발	국립재난안전연구원	2014.5. ~ 2014.8.	텍스트마이닝, 토픽분석	솔루션 도입
리얼상품평 프로젝트 언어처리 모듈 구축	GS홈쇼핑	2012. 2. ~ 2012. 8.	총 개발 책임	솔루션 도입