

# URL 기반 바이어 추천 시스템

품목분류 통합 DB 구축 및 바이어 맞춤 기업 추천모델 구현

나이스디앤비 X DIMA 3기 4조 1차 프로젝트

# INDEX

01

## 프로젝트 소개

- 프로젝트 주제
- 추진 배경 및 분석 목표
- 프로젝트 요약
- 분석 방법

02

## 프로젝트 진행 과정

- 데이터 수집
- 텍스트 전처리
- 품목분류 전처리
- 모델링

03

## 결론

- 프로젝트 기대 효과
- 2차 프로젝트 목표 및 계획

# 프로젝트 주제

## 주제

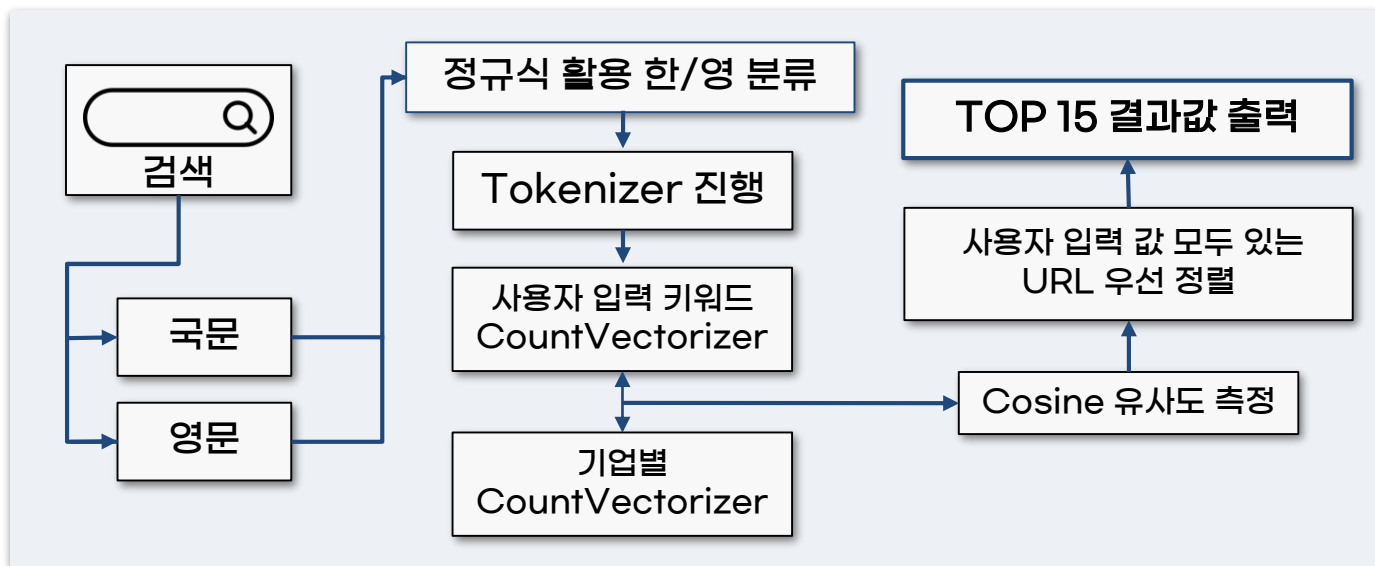
1. URL 및 품목 데이터를 구축
2. 데이터 기반으로 한 바이어 추천 시스템

## 타겟 고객

1. 해외 바이어
2. 수출을 희망하는 국내기업

# 프로젝트 요약

## 추천 시스템 모델 개발



## 분석 방법



국내 기업정보 DB | 사업자번호, URL 34만건



품목분류 데이터 (조달청 코드 기반) | 21만건



Tradekorea 홈페이지 카테고리표 (영문 품목 키워드)

# 분석 방법 / 데이터 분석 작업 프로세스

## 사전작업

## 크롤링

기업 URL 전처리



응답코드 처리



유효 URL 웹 크롤링 및 텍스트 추출

전처리



국문, 영문 키워드 카테고리 분류 및 전처리



기업 URL DB 전처리



품목분류 DB 구축



국내기업 DB 및 품목 DB 취합



추천 알고리즘 구현

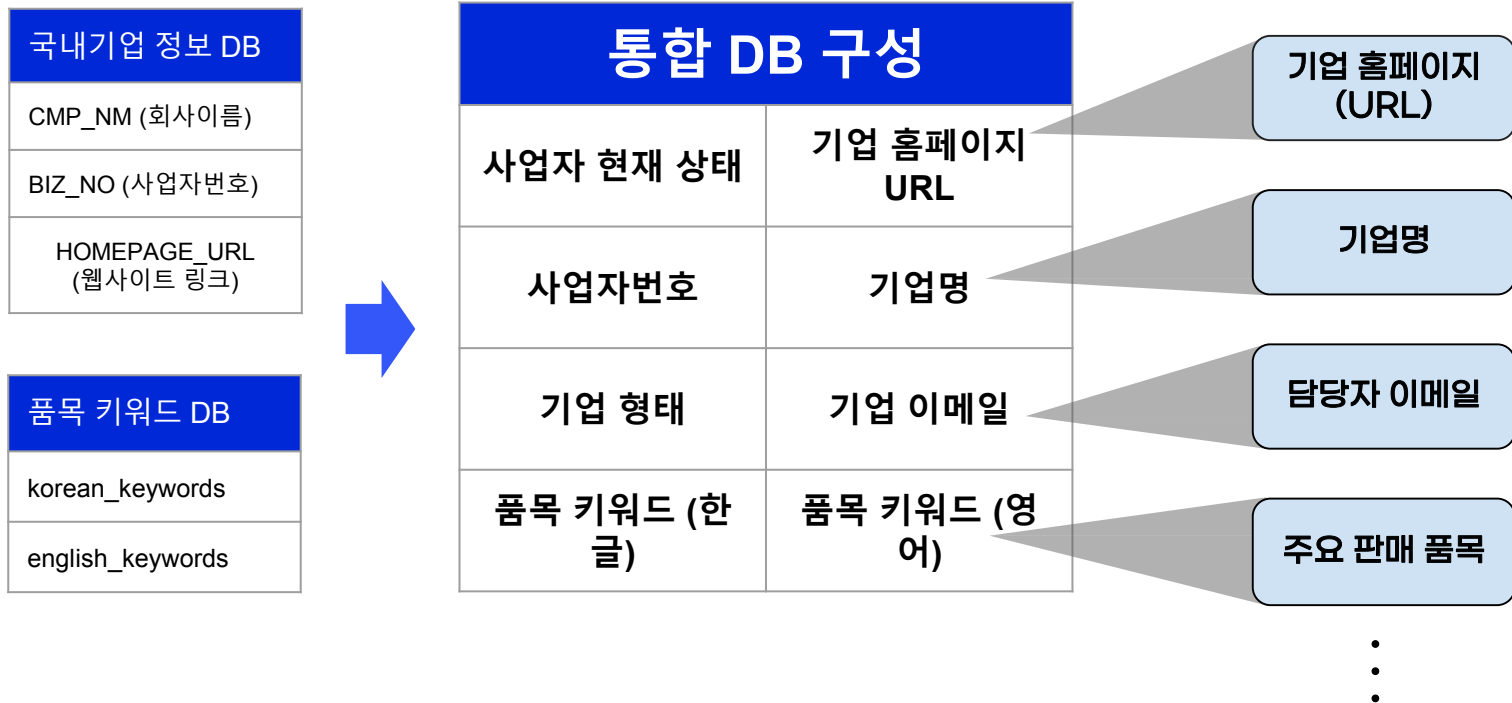


B2B 고객 대상 웹 서비스 구현

02

# 프로젝트 진행 과정

# 데이터 수집 / 필요한 데이터 요건 정의 및 수집





# 데이터 수집 / 기업 품목 데이터수집 방식의 근거

BIZ_NO	CMP_NM	HOMEPAGE_URL
1078841159	한우리해외사업본부	http://www.hanuri.edu.vn
1283990440	투리	http://www.vecka.co.kr
		stlive.co.kr
		eamse.com
		ungnail.com
2148804262	평로	http://www.pr701.com
5041786416	엘디티	http://www.luxurydol.com
1258134185	덕일힐타산업	http://www.dukil21.co.kr

나이스디앤비로부터 제공받은  
국내기업 URL DB

국내기업 URL DB를 활용하여  
주요 품목 정보와 이메일을  
수집해야 함



우리가 선택한 방식 = 크롤링

01

크롤링이란?

웹페이지를 그대로 가져와서 정보를 탐색하고 데이터를 추출해 내는 행위

02

크롤링 대상

나이스디앤비로부터 제공받은 국내기업 URL DB

03

크롤링 목적

각 기업 홈페이지 웹크롤링을 통해 텍스트를 수집하고 수집된 텍스트 안에서  
이메일과 품목 키워드를 추출하고자 함

# 데이터 수집 / 크롤링 작업방식 & <a>태그 수집 근거

<a> tag : HTML에서 'anchor'의 약자로 하이퍼링크를 생성하는 데 사용되는 태그

```
▼<li>
  <a href="/company/ceo.asp">회사소개</a>
</li>
▼<li>
  <a href="/business/package.asp">사업소개</a>
</li>
▼<li>
  <a href="/goods/pack_cap.asp">제품소개</a>
</li>
▼<li>
  <a href="/recruit/info.asp">인재채용</a>
</li>
```

```
▼<li>
  <a href="/goods/pack_cap.asp">제품소개</a>
  ▼<ul class="depth s3"> == $0
    ▼<li>
      <a href="/goods/pack_cap.asp">패키지 : 캡</a>
    </li>
    ▼<li>
      <a href="/goods/pack_bottle.asp">패키지 : 병</a>
    </li>
    ▼<li>
      <a href="/goods/pack_mold.asp">패키지 : 인몰드</a>
    </li>
    ▼<li>
      <a href="/goods/kitchen.asp">주방용품</a>
    </li>
    ▼<li>
      <a href="/goods/storage.asp">수납장</a>
    </li>
```

# 데이터 수집 / 1. url 사전 작업

URL 제거 기준	개수(건)
cafe.OOOO.OOO/	1,708건
blog.naver.com/	2,780건
smartstore.naver.com/	59,092건
www.instagram.com/	1,597건
한글도메인.kr/	6,100건
해당없음	203건

**삭제**

URL	개수(건)
기존 URL	347,684건

↓

URL	개수(건)
전체 URL	277,141건

# 데이터 수집 / 2. 모든 url에 http:// 추가

URL
www.example1.com
www.example2.com
http://www.example3.com
example4.com
http://www.example5.co.kr



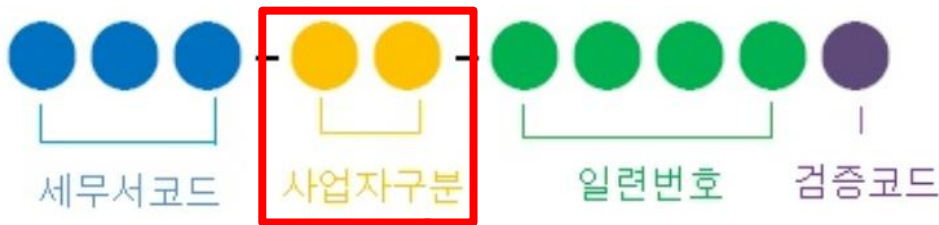
URL
<b>http://</b> www.example1.com
<b>http://</b> www.example2.com
http://www.example3.com
<b>http://</b> example4.com
http://www.example5.co.kr

# 데이터 수집 / 3. 응답코드 오류 사이트 제외

URL	STATUS CODE
http://www.example1.com	403
http://www.example2.com	200
http://www.example3.com	400
http://example4.com	200
http://www.example5.co.kr	200

URL	STATUS CODE
http://www.example2.com	200
http://example4.com	200
http://www.example5.co.kr	200

# 데이터 수집 / 4. 사업자등록번호 내 수출입과 관련없는 사업자 제외



**해당 URL  
삭제 처리**

- **80**    # 다단계 및 아파트관리사무소
- **82**    # 비영리법인
- **83**    # 국가,지방자치단체, 지방자치조합
- **84**    # 외국법인,연락사무소(대사관)
- **89**    # 종교단체

# 데이터 수집 / 5. 크롤링 - 텍스트

NICE dun & bradstreet

회사소개 서비스안내 투자정보 홍보센터 고객센터 인재채용

## 아시아 제일의 Global Commercial Credit Bureau

Create Further Knowledge  
Discover More Information  
Deliver Better Solutions

㈜NICE디앤비는 국내 제1의 금융인프라 그룹인 NICE그룹과  
200년 전통의 세계최대 기업정보기관인 D&B Corporation 등의 합작법인으로 2002년 10월 설립되었으며,  
2005년 금융위원회로부터 신용정보 조항 및 조사에 관한 국내 사업 라이선스를 취득한  
"국내 최대의 글로벌 기업"

**01 첫 페이지 데이터 수집**

보도자료

- ESG 만·권 활동지원단체 NICE디앤비 발족, 중소기업 ESG 경영 진단과 컨설팅 실시 ...
- NICE디앤비, 신용조사회업 혁신에 대한 공헌을 인정받아 2023년 혁신대상 수상

Dimensions Responsive 791 x 599 100% No throttling

Always match Chrome's language Switch DevTools to Korean Don't show again

Elements Console Sources Network

IS NICE 회사 소개 엔진오일 배터리 자동차 부품 자동차 액세서리 원단 컬러리

a.dropdown-toggle 89.84 x 84 #333333  
Color  
Font 14,2416px notok; "Apple SD Neo Gothic...  
Padding 32px 12px  
ACCESSIBILITY  
Contrast Aa 12.63  
Name 엔진오일 외  
Role link  
Keyboard-focusable

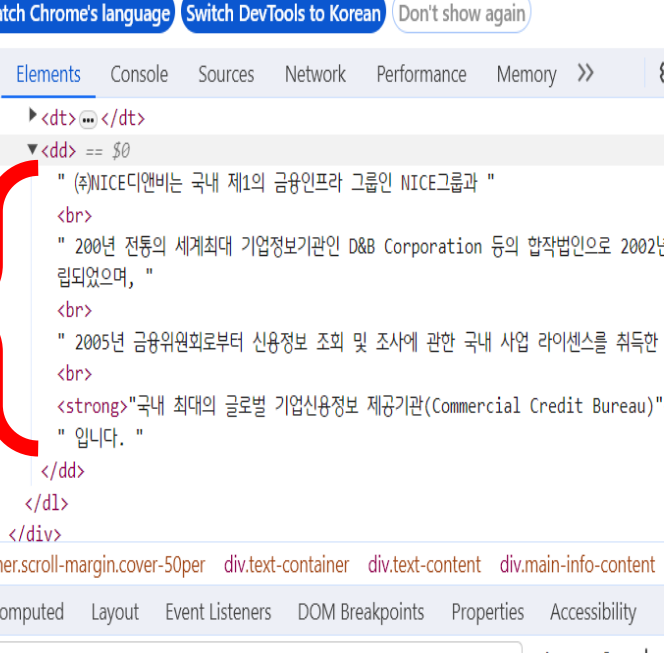
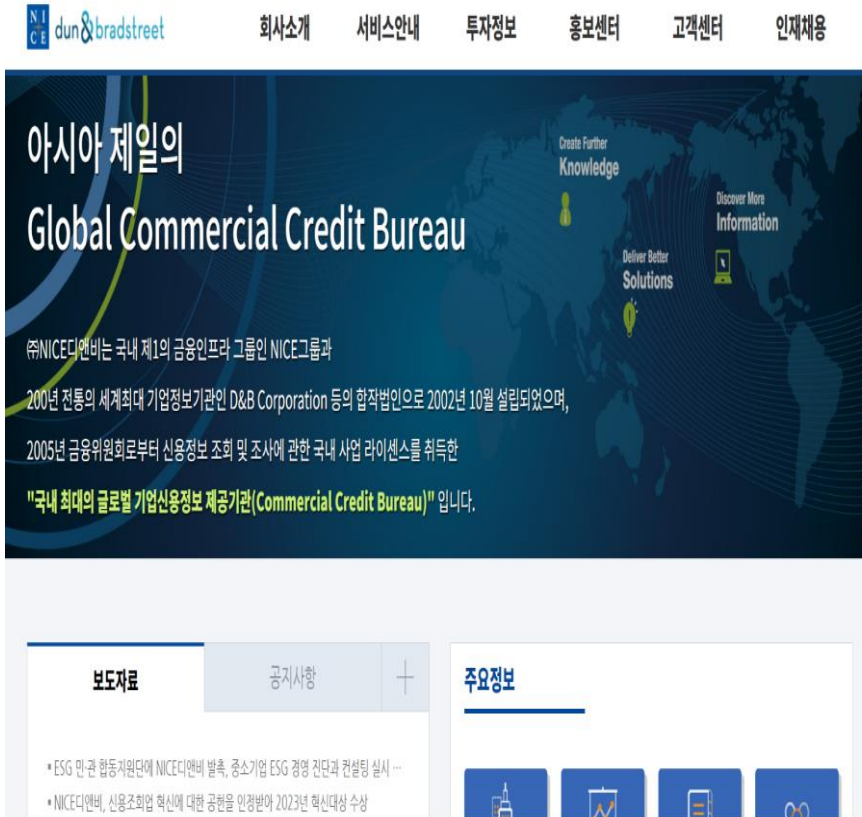
**02 <a tag> 30개의 데이터 수집**

```
<div>
  <div class="collapse navbar-collapse">
    ::before
    <ul class="nav navbar-nav navbar-right">
      ::before
      <li class="dropdown" id="gnb_55">
        <li class="dropdown" id="gnb_56">
          <a href="#" class="dropdown-toggle" data-toggle="dropdown" aria-expanded="false">엔진오일 외</a>
          <ul class="dropdown-menu">
            <li>
              <div class="collapse navbar-collapse">
                <ul class="nav navbar-nav navbar-right">
                  <li class="dropdown">
                    <a href="#" class="dropdown-toggle" data-toggle="dropdown" aria-expanded="false">배터리</a>
                    <ul class="dropdown-menu">
                      <li>
                        <a href="#" class="dropdown-toggle" data-toggle="dropdown" aria-expanded="false">자동차 부품</a>
                        <ul class="dropdown-menu">
                          <li>
                            <a href="#" class="dropdown-toggle" data-toggle="dropdown" aria-expanded="false">자동차 액세서리</a>
                            <ul class="dropdown-menu">
                              <li>
                                <a href="#" class="dropdown-toggle" data-toggle="dropdown" aria-expanded="false">원단</a>
                                <ul class="dropdown-menu">
                                  <li>
                                    <a href="#" class="dropdown-toggle" data-toggle="dropdown" aria-expanded="false">컬러리</a>
                                  </li>
                                </ul>
                              </li>
                            </ul>
                          </li>
                        </ul>
                      </li>
                    </ul>
                  </li>
                </ul>
              </div>
            </li>
          </ul>
        </li>
      </li>
    </ul>
  </div>
</div>
```

엔진오일 / 배터리 / 자동차 액세서리 / 섬유원단

Uncaught TypeError: Cannot read properties of null (reading 'style')  
at s (roughmap.ender.js:2:15392)  
at RoughMapGenerator.generate (roughmap.ender.js:2:18921)  
at daum.roughmap.onDataLoad (roughmap.ender.js:2:16493)  
at x6rj.1son:1:115

## 첫 페이지 내 모든 텍스트 데이터 수집



DevTools is now available in Korean!

Always match Chrome's language Switch DevTools to Korean Don't show again

Elements Console Sources Network Performance Memory >>

<dt> </dt>

<dd> == \$0

" (주)NICE디앤비는 국내 제1의 금융인프라 그룹인 NICE그룹과 "

<br>

" 200년 전통의 세계최대 기업정보기관인 D&B Corporation 등의 합작법인으로 2002년 10월 설립되었으며, "

<br>

" 2005년 금융위원회로부터 신용정보 조회 및 조사에 관한 국내 사업 라이선스를 취득한 "

<br>

<strong>"국내 최대의 글로벌 기업신용정보 제공기관(Commercial Credit Bureau)"</strong>

" 입니다. "

</dd>

</dl>

</div>

<div>container.scroll-margin.cover-50per div.text-container div.text-content div.main-info-content dl dd

Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessibility

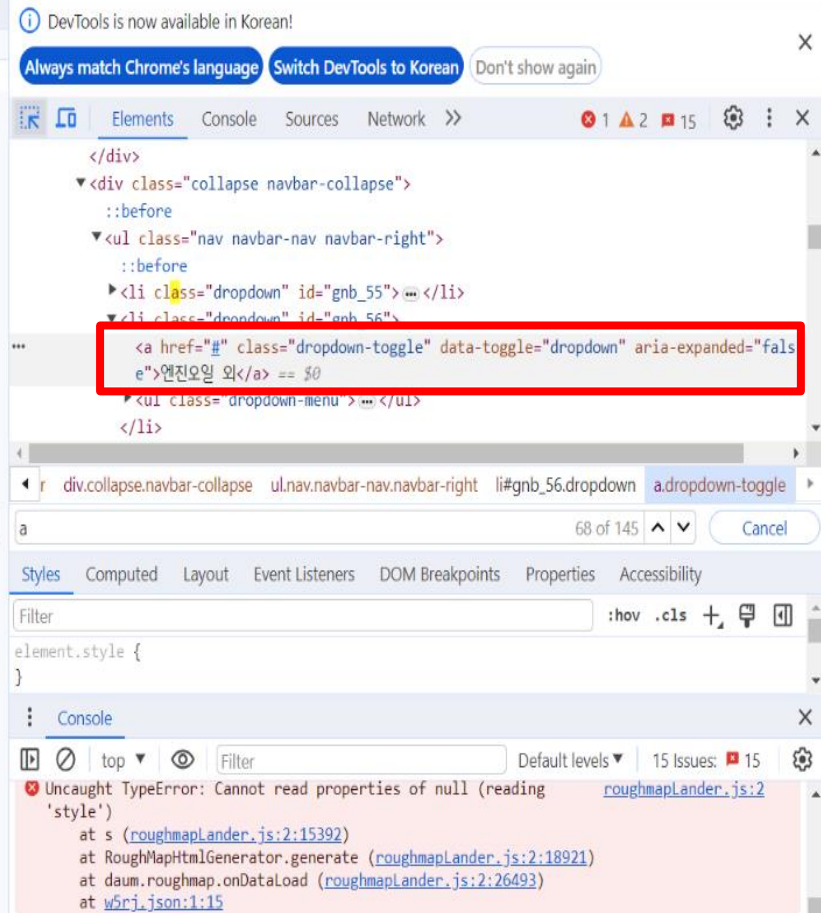
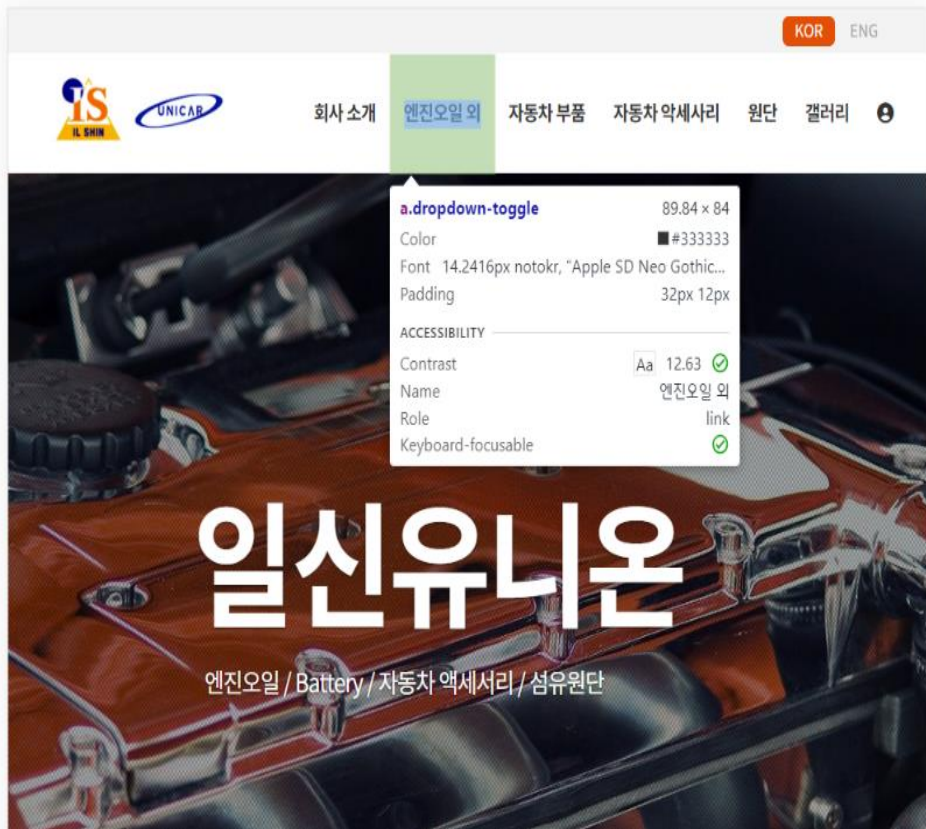
Filter :hov .cls + -

element.style {

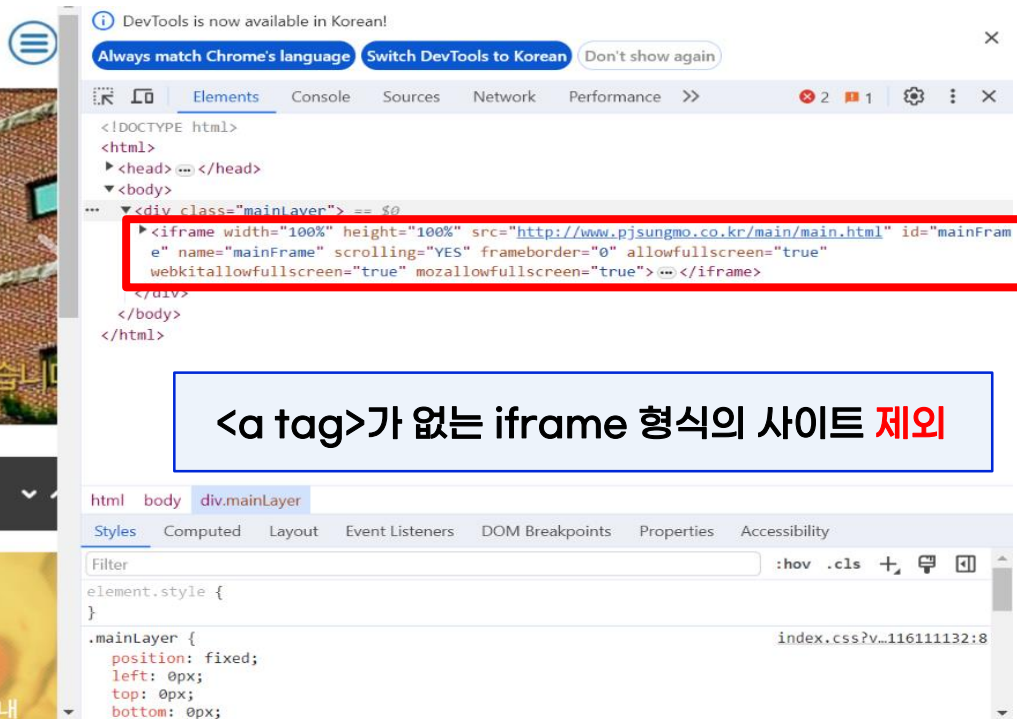
}



## &lt;a tag&gt; 30개의 텍스트 데이터 수집



# 데이터 수집 / 5. 크롤링 - 제외 사이트



<a tag>가 없는 iframe 형식의 사이트 제외

# 데이터 수집 / 5. 크롤링 - 제외 사이트



# 데이터 수집 / 5. 크롤링 - URL DB 보완



사업자 현재상태 확인  
※국세청 홈택스 실시간 정보제공

✓ **계속사업자**  
✗ 폐업자  
✗ 휴업자

비즈노사이트 크롤링을 통해 URL DB 보완

# 데이터 수집 / 6. 크롤링 완료한 URL DB

사업자등록번호	URL	TEXT
8232401301	http://www.hankook-e.co.kr	['초과한 데이터는 12월 12일에 고지되어 있습니다. 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
2630202393	http://www.mapsip.com/front/gnu/index2.php	[' >₩n ₩nKorean₩n English₩n Japanese₩n Translating 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
1454500492	http://www.leadspiano.com	[' L-850 유압 (높,낮이 조절) ₩n운동방법: 유압 조절(매우 간편함) ₩n L-152 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
7118802445	http://www.bi5882.com	['로그인₩n LME 시세 (Ton/\$) ₩n더보기 +₩n날 짜 ₩n 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
8498802319	http://www.sutech.kr	['회사소개₩n비즈니스₩n제품소개₩n고객센터₩n상담문 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
1798102400	http://www.protechems.com	['', 'We are excited to announce that the IIS.NET Forum 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
5708702167	http://www.usgne.go.kr	"교육환경정보시스템 ₩n알림 ₩n메뉴 ₩n일반인 ₩n학교관 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
1268683383	http://www.shinyu.kr	['031-559-1119 ₩nInnovation ₩n고객만족을 위한 끝없는 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
2958102376	http://www.ilstudio.co.kr	['.cls-1{fill:#777;}.cls-2{isolation:isolate;}.cls-3{fill:#175fc4; 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
8302701085	http://www.eylpartners.com	['EMAIL ₩nCALL ₩nMAP ₩nCompany ₩nProduct ₩nNews ₩nSupport ₩nKR', 'E 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
3798102335	http://www.elpisline.com	['팝업레이어 알림이 없습니다. ₩n메뉴 열기 ₩n국제운송 전문서비스 기업 ₩nEG 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
1074474296	http://www.dongrae.nonghyupi.com	['SULinux APM_Install-tools에 의해 ₩nApache, MySQL, PHP 설치 되었습니다. 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
3078517184	http://kosee.jams.or.kr	['(사)한국환경교육학회 ₩n학회소개 ₩n논문검색 ₩n정보마당 ₩n공지사항 ₩n더 계속 사업자 ※ 국세청 홈택스 실시간 정보제공
646802360	http://www.YKtower.co.kr	['홈 로그인 회원가입 사이트맵 contact us ₩n ' , '홈 로그인 회원가입 사이트 계속 사업자 ※ 국세청 홈택스 실시간 정보제공

## 각 회사별 텍스트 추출완료

## 텍스트에서 추출할 주요 데이터

### 1. 기업 이메일 정보

### 2. 기업 주요 품목 정보



# 텍스트 전처리 / 1. 이메일 수집

## 추출된 텍스트

\nProduct\nNews Release\nbusiness@gmail.com  
Overview\nAccess\nInquiry\nCreating the future  
environment with exhaust gas treatment technology\n

## 이메일 정규식 사용

```
r'\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Z|a-z]{2,}\b'
```

## 추출된 텍스트에서

business@gmail.com

수집

# 텍스트 전처리 / 3. 한/영 텍스트 추출

(繁體)\n中文 (简体)\n製品情報\n新着情報\n企業概要\n採用情報\nアクセス\nお問い合わせ\n排ガス処理技術で  
未来環境を創造する\n



한자 및  
기타 외국어  
텍스트 제외

\nProduct\nNews Release\nbusiness@gmail.com  
Overview\nAccess\nInquiry\nCreating the future  
environment with exhaust gas treatment technology\n



영문 텍스트  
추출

\n제품정보\n새소식\n기업개요\n오시는 길\n문의\n배기가  
스 처리 기술로 미래 환경을 창조하겠습니다.\n



한글 텍스트  
추출

# 텍스트 전처리 / 4. 추출한 텍스트 토큰화

\nProduct\nNew Release\nBusiness  
Overview

'product', 'new', 'release', 'business', 'overview',  
'access', 'inquiry', 'future', 'environment', 'gas', 'treatment',  
'technology'

**nltk 내 word/sentence tokenize 모듈  
및 pos\_tag, WordnetLemmatizer 활용**

\n제품정보\n새소식\n기업개요\n오시는 길\n문의  
다.

'제품정보', '새소식', '기업개요', '오시는 길', '문의', '배기가스',  
'처리', '기술', '미래', '환경', '창조'

**Mecab & nltk.pos 함수 활용**



# 텍스트 전처리 / 4. 추출한 텍스트 토큰화의 결과

url	email	korean	english
http://www.marryfrill.com	marryberry1@naver.com	예금주, 사업자, 정보, 확인, 예금 주, 사업자, 정보, 확인, 아이디, 가입, 방...	home, shoe, acc, lookbook, press, board, membe...
http://www.sendbook.co.kr	helpcustomer@naver.com	쇼핑, 행복, 사람, 추천, 교육, 공 식, 파트너, 관심, 고객, 도움, 검 색어, ...	version, mbn, wjdd, rlaa, friend, friend, biz,...
http://www.luremarket.kr	luremarket@naver.com	즐거찾기, 추가, 회원, 입시, 포인 트, 혜택, 로그인, 회원, 가입, 장 바구니, ...	sale, product, notice, review, arrival, j, rxe...
http://www.untagestore.com	nuntage@naver.com, untage@naver.com	공지, 사항, 공지, 사항, 단독, 오 픈, 전체, 제목, 내용, 글쓴이, 아 이디, ...	shop, information, community, sign, order, bag...
http://www.joystory.shop	njoystorenn@mail.ru	회원, 가입, 로그인, 등록, 음악, 동영상 커뮤니티 미니 게임 추...	joy, joy, tilda, tilda, pricing, madeontilda, ...
http://www.goodfarm.net/daesan			daesan

**각 회사별 주요품목 추출완료**  
( 22만건 -> 5만 8천건 )

# 품목분류 전처리 / 5. 품목분류 키워드 생성 - 한글

음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 로얄스위트 멜론
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 살린 멜론
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 스타스트라이프 멜
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 슈가베이비 수박
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 타이거베이비 멜론
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 옐로베이비 수박

21만여건의 품목 데이터에서  
약 22000개 키워드 추출

문제점

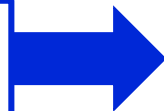
- 불필요한 고유명사 (EX: 사람 이름)
- 지나치게 포괄적인 단어
- 품목 대신 특정 상품이 나오는 이상치

# 품목분류 전처리 / 5. 품목분류 키워드 생성 - 한글

음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 로얄스위트 멜론	
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 로얄스위트 멜론	비유전자조작) 건조 과일>비유전자조작) 건조 멜론
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 로얄스위트 멜론	비유전자조작) 건조 과일>비유전자조작) 건조 멜론
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 로얄스위트 멜론	비유전자조작) 건조 과일>비유전자조작) 건조 멜론
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 로얄스위트 멜론	비유전자조작) 건조 과일>비유전자조작) 건조 멜론
음식료품 및 담배제품>비유전자조작) 건조 과일>비유전자조작) 건조 멜론>비유전자조작) 건조 로얄스위트 멜론	비유전자조작) 건조 과일>비유전자조작) 건조 멜론

‘건조’, ‘과일’, ‘멜론’

불필요 문자, 중복값 제거 후 전처리 진행



약 6000개  
키워드 생성

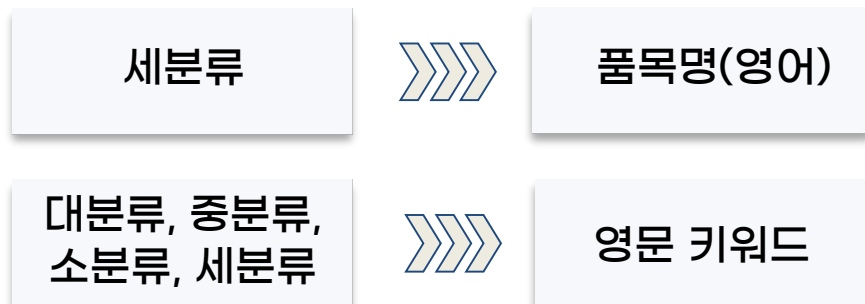
# 품목분류 전처리 / 5. 품목분류 키워드 생성 - 영문

## Category



출처: 트레이드코리아 홈페이지

## 영문 품목 분류



토큰화, 불용어 제거 및  
표제어 추출 진행

# 품목분류 전처리 / 6. Bag of Words 구성

**최종 DB  
구축 완료**

url	가공품구금넷노능닥리늄엔진자동차신스히알루론산히어로치터트팅지링																		
http://www.unicar.co.kr	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0
http://www.sunghwamold.com/	0	0	0	0	0	0	35	0	0	0	10	0	0	0	0	0	0	0	0
http://www.gs-one.co.kr	0	0	0	0	0	0	0	0	0	1	15	0	0	0	0	0	0	0	0

3 rows x 6868 columns

한글 품목 키워드  
6,868건

url	ab abrasive absorber ac access accessory acdc acid acoustic acrylic ... wrapping wrench wristb																		
http://www.atm-studio.com	0	0	0	0	2	0	0	0	0	0	...	0	0	...	0	0	0	0	0
http://www.storekhan.com	0	0	0	0	0	0	0	0	0	0	...	0	0	...	0	0	0	0	0
http://www.kissel.co.kr	0	0	0	0	0	0	0	0	0	0	...	0	0	...	0	0	0	0	0

3 rows x 1428 columns

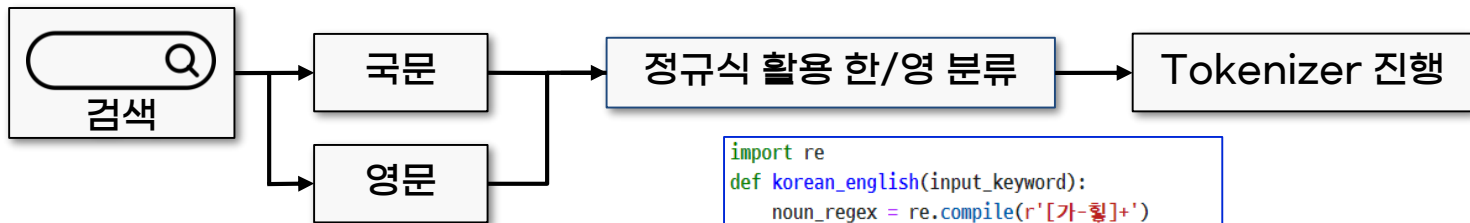
영어 품목 키워드  
1,428건

# 모델링

---

- 사용자가 입력한 값에 대해 토큰화와 Countvectorizer 진행
- 이후 추출된 키워드와 BoW 데이터 간 Cosine 유사도 측정
- 유사도 순 정렬 후 추출된 키워드를 모두 포함하는 URL을 우선순위로 출력

# 입력값 토큰화



```

while(1):
    input_keyword = input('키워드 입력: ')
    korea_english = korean_english(input_keyword)
    if korean_english(input_keyword) == 1:
        input_keyword = tokenizer(input_keyword)

    else: # 영어 버전
        input_keyword = input_keyword.lower()

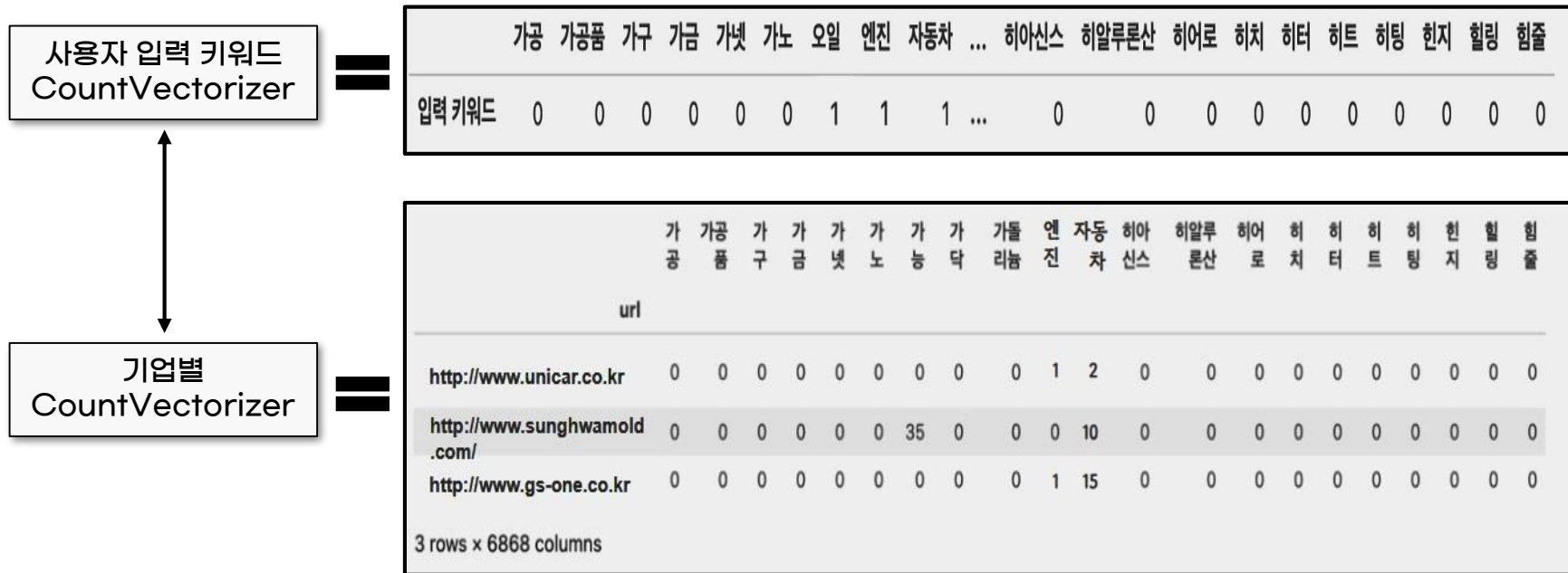
        # 불용어 제거 및 단어만 추출
        stopwords = nltk.corpus.stopwords.words('english')
        all_tokens = []

        word_tokens = word_tokenize(input_keyword)
  
```

```

import re
def korean_english(input_keyword):
    noun_regex = re.compile(r'[가-힣]+')
    nouns = noun_regex.findall(input_keyword)
    if nouns: # 한글이 존재 한다면
        return 1 # 한글 존재 1반환
    else:
        return -1 # 영어 존재
  
```

# 토큰화된 단어의 CountVector화





# 유사도 측정 및 결과값 출력

Cosine 유사도 측정

사용자 입력 값을  
모두 포함하는 URL 우선 정렬

TOP 15 결과값 출력

	url	CMP_NM	BIZ_NO	email	similarity	status
1위	http://www.unicar.co.kr	일신유니온	1168126245	unicar@unicar.co.kr, nunicar@unicar.co.kr	0.63	계속 사업자
2위	http://www.sunghwamold.com/	성화금형	6068636969	nkim2525@hanmail.net, nshcast1@naver.com	0.58	계속 사업자
3위	http://www.gs-one.co.kr	지에스원	1288616158	globalsimulatorone@gmail.com	0.48	계속 사업자
4위	http://www.kkengine.com	금강산업엔진	4188118578	info@kkengine.co.kr	0.46	계속 사업자
5위	http://www.xado.co.kr	한국기술무역	1208602951	nwww.jongnogi@hanmail.net	0.43	계속 사업자

# WHY 코사인유사도

## Word2Vec

```
keyword=['자동차']
similar_words = model.wv.most_similar(keyword)
print(similar_words)
```

```
[('레사', 0.35070157051086426), ('호르몬', 0.3380924165248871), ('프라도', 0.3296448290348053), ('유선', 0.3290241062641144), ('포에버', 0.31563621759414673), ('응집', 0.31146612763404846), ('케이', 0.31046363711357117), ('기상', 0.30747318267822266), ('후고', 0.3029279112815857), ('증류기', 0.29536792635917664)]
```

원하는 결과 값 도출 어려움

```
keyword=['기능성', '제품']
similar_words = model.wv.most_similar(keyword)
print(similar_words)
```

```
-----
KeyError                                Traceback (most recent call last)
Cell In[35], line 2
      1 keyword=['기능성', '제품']
----> 2 similar_words = model.wv.most_similar(keyword)
      3 print(similar_words)
```

학습데이터 이외의 키워드를 입력할 경우 오류 발생

# WHY 코사인유사도

SequenceMatcher

입력값 : '자동차', '엔진', '오일'

	url	CMP_NM	BIZ_NO	email
1위	http://www.beforever.co.kr	나눔정보	7858101303	beforever@daum.net
2위	http://www.adellife.com	아델라이프	6081418956	adellifer@gmail.com
3위	http://www.farmsale.kr	마노기술	1838101074	0707507@naver.com, parkingsale@daum.net
4위	http://ysgpc.modoo.at/	연세거평의원	7799000510	yonseipcr@naver.com
5위	http://www.gtli.co.kr	지티엘파크	2208770023	gtl1004@gmail.com
6위	http://www.flapborn.com	플랩본	2068800853	helpcustomer@naver.com
7위	http://www.marketingmonster.co.kr	마케팅몬스터	8571200734	nananom@nate.com, ananom@nate.com

낮은 정확도 &amp; 결과 값 출력 시간이 오래 걸림

# Cosine 유사도 확인



사용자 입력값이 주요 품목인 URL 상위 링크

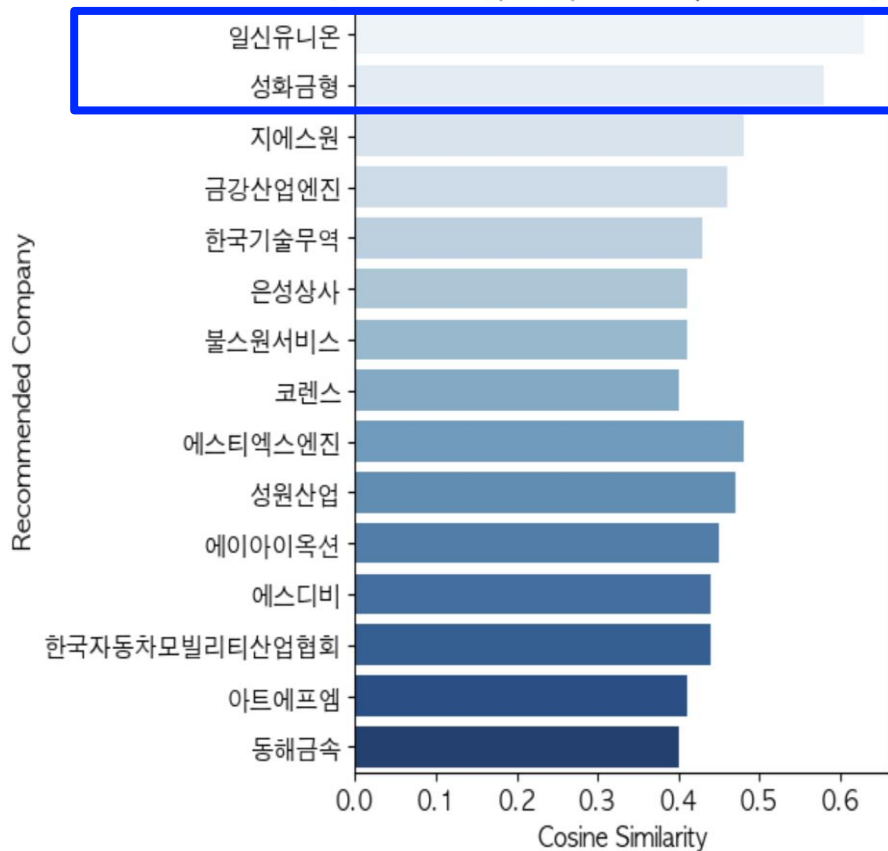
Cosine 유사도 높은 순으로 TOP 15위 출력

# Top 15 추천 URL

	url	CMP_NM	BIZ_NO	email	similarity
1위	http://www.unicar.co.kr	일산유니온	1168126245	unicar@unicar.co.kr, nunicar@unicar.co.kr	0.63
2위	http://www.sunghwamold.com/	성화금형	6068636969	nkim2525@hanmail.net, nshcast1@naver.com	0.58
3위	http://www.gs-one.co.kr	지에스원	1288616158	globalsimulatorone@gmail.com	0.48
4위	http://www.kkengine.com	금강산업엔진	4188118578		0.46
5위	http://www.xado.co.kr	한국기술무역	1208602951	nwww.jongnogi@hanmail.net	0.43
6위	http://www.esmotor.com	은성상사	5142664124		0.41
7위	http://bullsone.com	볼스원서비스	1208776635		0.41
8위	http://www.korens.co.kr	코렌스	6218122590	recruit@korens.com	0.40
9위	http://www.stxengine.co.kr	에스티엑스엔진	6098162781	n055-280-0835sympark@stxengine.co.kr, nswjeong@...	0.48
10위	http://www.swindustry.net	성원산업	3018184028		0.47
11위	http://aiauction.co.kr	에이아이옥션	2548100931	nadmin@kengine.io, admin@aiauction.co.kr	0.45
12위	http://www.sdbbellows.com	에스디비	3118115232	sdb@sdbbellows.com	0.44
13위	http://kama.or.kr	한국자동차모빌리티산업협회	1168203201		0.44
14위	http://www.artfminc.com	아트에프엠	1198178548	artfminc@korea.com	0.41
15위	http://www.dhmetal21.com	동해금속	5051792828		0.40

코사인유사도,  
키워드가  
모두  
있는 것부터  
상위 정렬

키워드 "자동차, 엔진, 오일" Top 15 코사인 유사도



03

결론

# 프로젝트 기대 효과



추천 모델을 통한  
프로세스 간소화로 인한 시간 단축