



HyTV

Hynix Trend Viewer

We Do Trend

Data Analytics Technology
Mobile & Chatbot
Software & Intelligence
Data Analytics Technology

박헌준
김다혜
이수연
채선율

I Motivation

II Introduction

III About HyTV

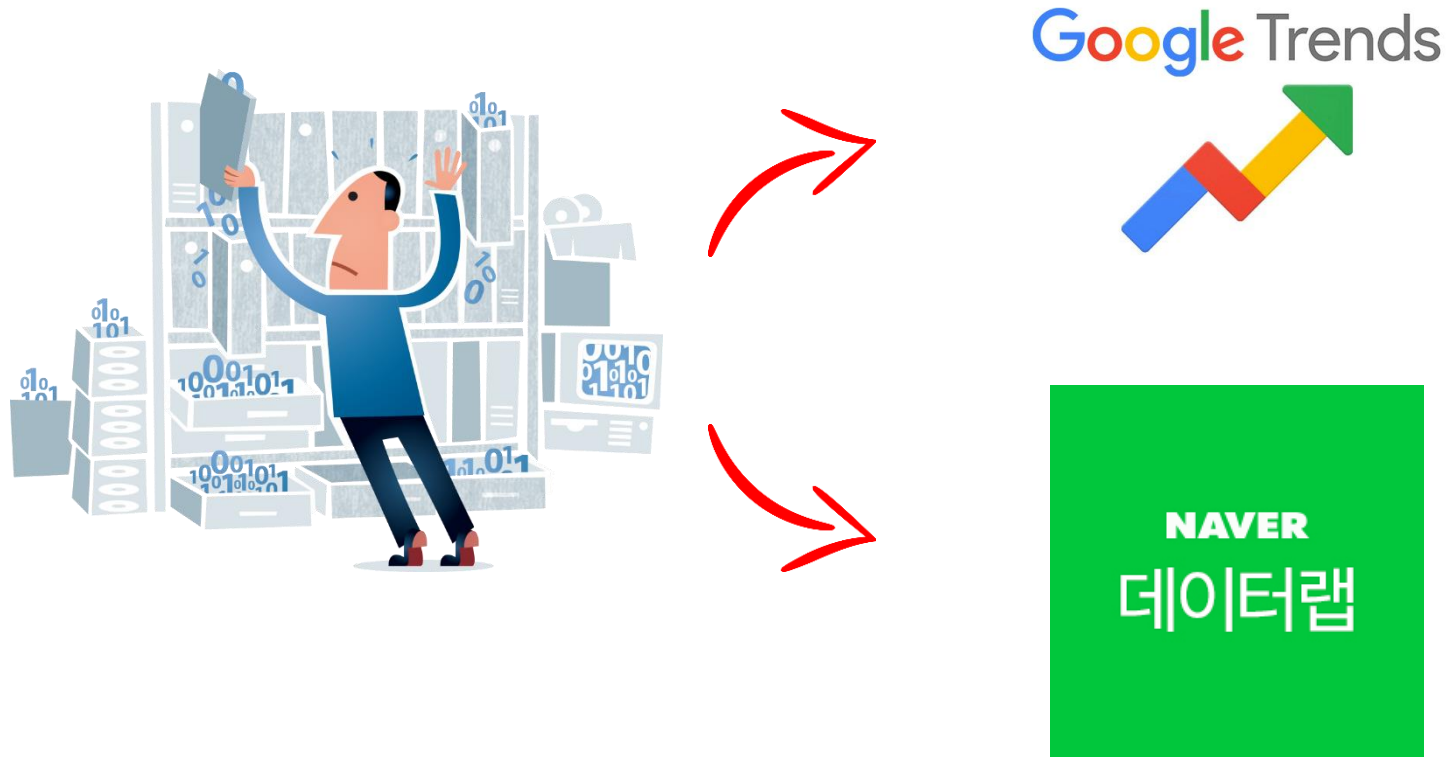
1. Data Collection & Data Preprocessing
2. Topic Modeling
3. Sentimental Analysis
4. Web Interface

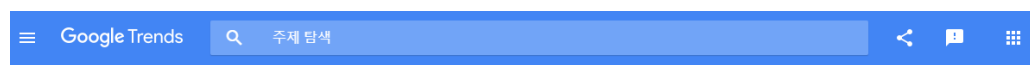
IV Demonstration

V Summary

Motivation

빅데이터 시대





2018년 인기 검색어 보기 - 대한민국

종합

- 1 로스트아크
- 2 월드컵
- 3 김비서가 왜 그럴까
- 4 외모지상주의
- 5 평창올림픽

인물

- 1 조민기
- 2 구하라
- 3 노회찬
- 4 판빙빙
- 5 김보름

국내 뉴스/이슈

- 1 비트코인
- 2 태풍 경로
- 3 논산 여교사
- 4 가상화폐
- 5 드루킹

영화

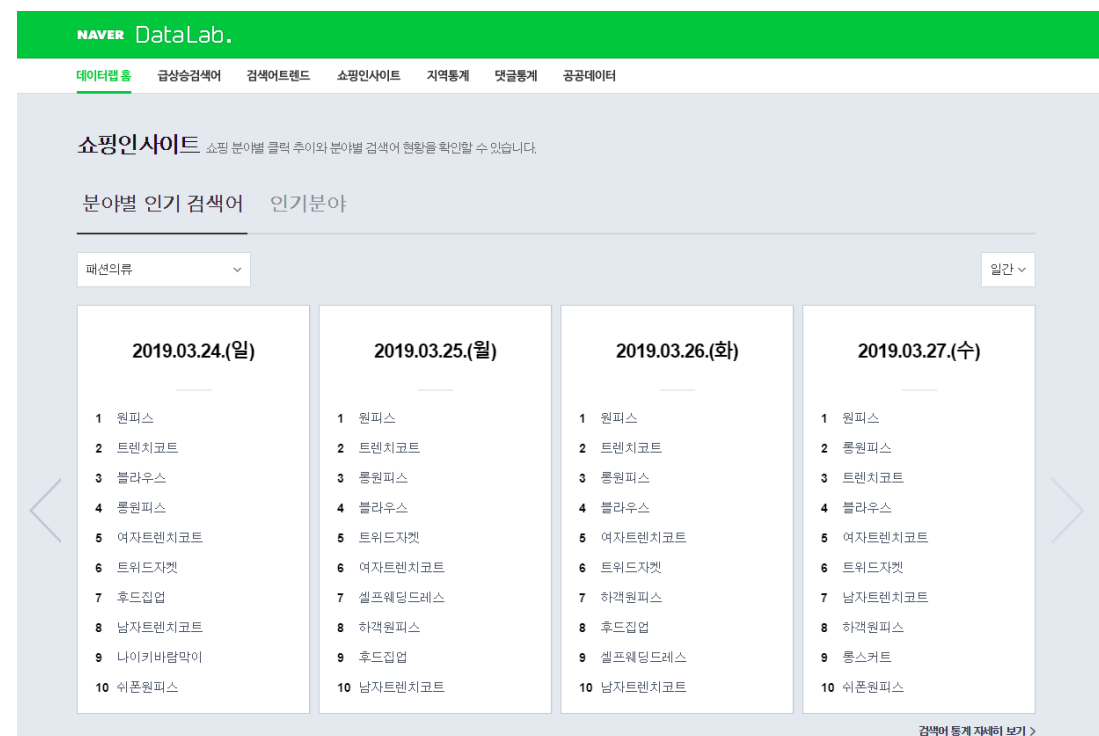
- 1 신과함께
- 2 마녀
- 3 블랙 팬서
- 4 보헤미안 랍소디

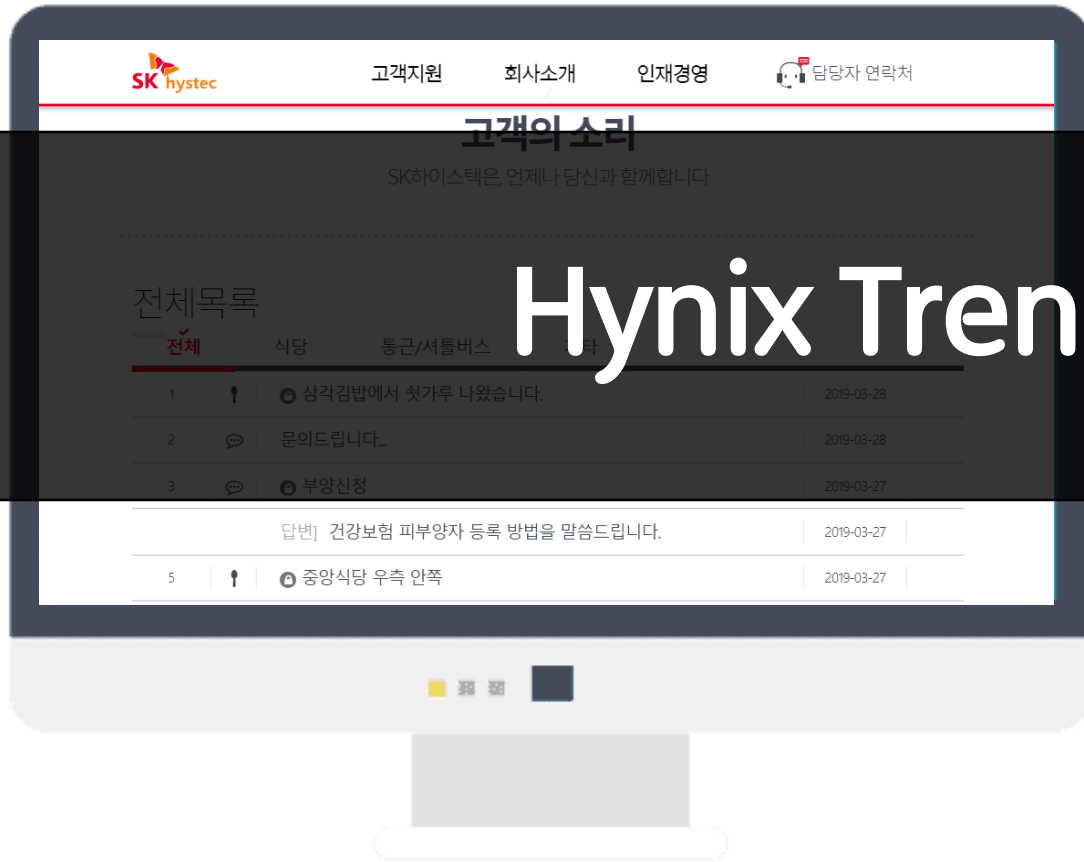
TV 프로그램

- 1 김비서가 왜 그럴까
- 2 하트시그널 2
- 3 프로듀스 48
- 4 슬기로운 감빵생활

게임

- 1 로스트아크
- 2 소울워커
- 3 듀랑고
- 4 메이플스토리





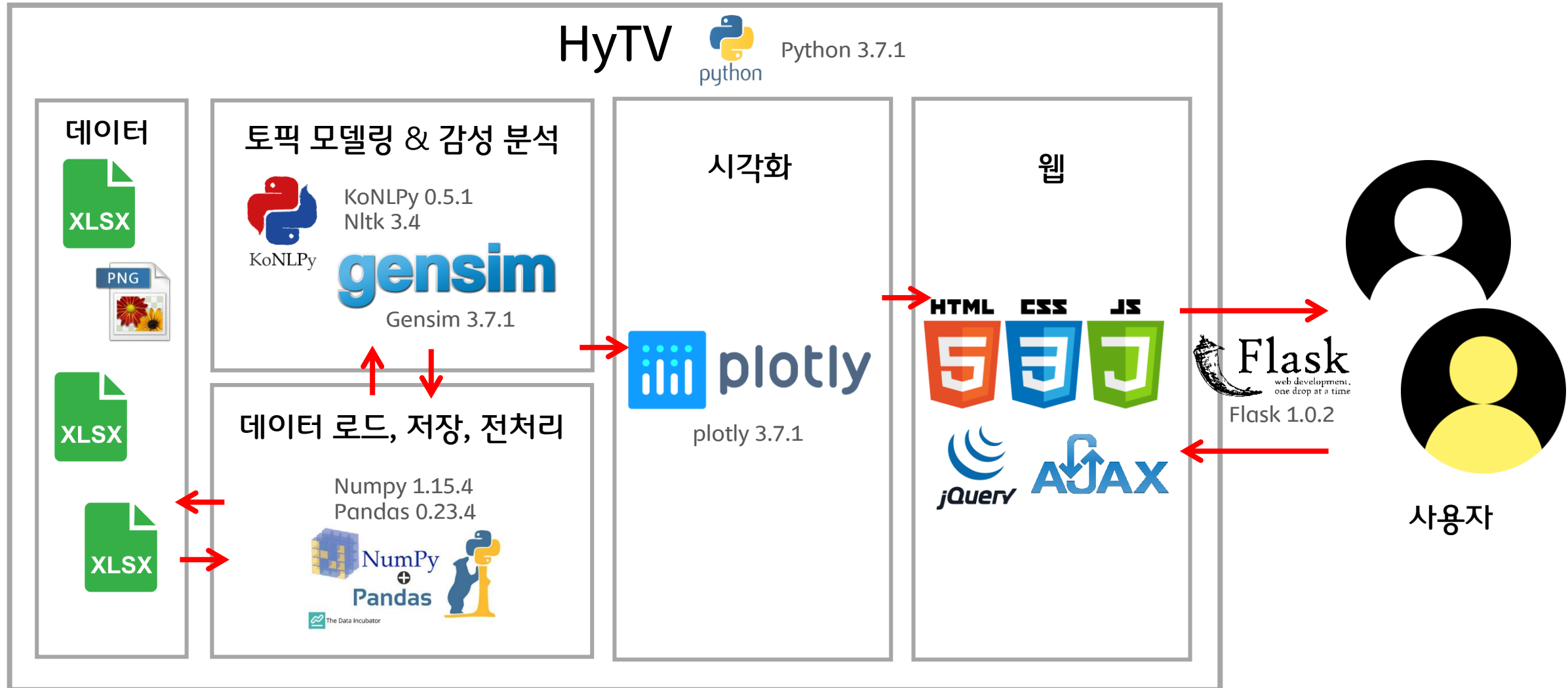
Hynix Trend Viewer

식당 관련 글만
보고 싶은데..

그래서 요즘
문제가 뭐야?

Introduction

개발 환경 및 프레임워크



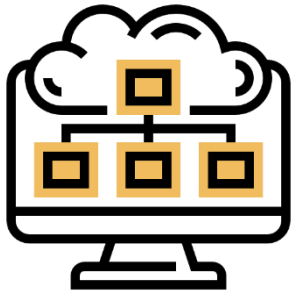
About HyTV

Work Flow

1. Data Collection
& Preprocessing

데이터 파싱

2. Topic Modeling



토픽 모델링

3. Sentiment
Analysis

게시물 긍 · 부정

4. Web Reporting



HyTV 인터페이스

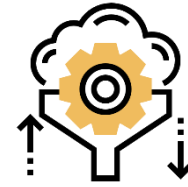
1. Data Collection & Preprocessing



Raw Data



Crawling



Parsing

| 전체 | 식당 | 통근/서플버스 | 기타 |
|----|----|------------------------------|------------|
| 1 | ! | ! 상감김밥에서 썰가루 나왔습니다. | 2019-03-28 |
| 2 | ! | 문의드립니다.. | 2019-03-28 |
| 3 | ! | ! 부양신청 | 2019-03-27 |
| | | 답변] 건강보험 피부양자 등록 방법을 말씀드립니다. | 2019-03-27 |
| 5 | ! | ! 중앙식당 우측 안쪽 | 2019-03-27 |
| 6 | ! | ! 버스 분실물 | 2019-03-27 |
| | | 답변] 버스 분실물 | 2019-03-27 |
| 8 | ! | ! 건강식단 | 2019-03-27 |
| 9 | ! | ! 보안검색 근무자 관련 문의합니다. | 2019-03-26 |
| | | 답변] 보안검색 근무자 관련 문의합니다. | 2019-03-26 |

| datetime | title | service | voc_type | voc_date | text | |
|---------------------|-----------------------|---------------|--------------|------------|----------------------------------|---------------------------------|
| 2019-03-21 오전 11:51 | 서울 작석이 고장났습니다 > | | | | 작곡 동료쪽 6번패출 작석이 고정되어있어 등반이 | |
| 2019-03-21 오전 8:12 | 청운 4관 2중에 사는데 불면! > | 이전 > 기사 | 불면 / 불면 | 2019-03-21 | 설날 전에 떨어진 화장실 물바누는 다시 채워질 생 | |
| 2019-03-20 오후 11:17 | 편의식 선택 > | 식당 > 이전 > 복지 | 편의 / 복지 | 2019-03-20 | 편의식 건의드립니다.건강식도 좋지만 우선 배가 안 | |
| 2019-03-20 오후 6:05 | 통신사 KT > | 기타 > 청주 > 기타 | 불면 / 불면 | 2019-03-20 | 통신사 KT를 사용중인데 너무 안터져서 확인해보 | |
| 2019-03-20 오후 12:56 | 기분 나빠요 > | | | | 아침부터 기분이 안좋네요테크아웃 바뀌고 좋아 | |
| 2019-03-19 오전 10:07 | 건강식 1번 샐러드(단품) > | 이전 > 복지 | 불면 - 기타 | 2019-03-19 | 요즘에 건강식으로 샐러드 나오는데 1번 샐러드(단 | |
| 2019-03-18 오전 8:16 | 다이어트 메뉴 부활 요청 > | 식당 > 이전 > SKY | 건의 | 2019-03-15 | 갑자기 다이어트 메뉴가 없어졌네요. 없어진다고 공 | |
| 2019-03-17 오후 11:13 | 제발요 히터 좀 꺼주세요 > | | | | 아직도 기사님은 겨울인가? 혼자 셔츠만 입고 계신 | |
| 2019-03-17 오후 5:14 | 식단에서 꼭 일부러 사용이 > | | | | 가라게요요아이스 맨즈가조요아이스음식이 | |
| 2019-03-16 오후 10:16 | 행복마을 정갈... > | 식당 > 이전 > 행복 | 1 칭찬 - 맛&품질 | 2019-03-15 | 요새 행복마을 너무 맛있어졌 것 같아요~~??? | |
| 2019-03-15 오후 9:08 | 오픈마켓 오늘도 안나왔네요 > | 식당 > 청주 > 1 | 행퍼 문의 | 2018-03-15 | 이번주에는 오픈마켓 나옴을 알았는데 나오지가 않 | |
| 2019-03-15 오후 6:09 | 동탄 5시 퇴근버스 노선 속가 > | | | | 4시 55분에 출발한 동탄 5시버스가 우남 퍼스트빌 | |
| 2019-03-15 오후 12:11 | 위생교육이 필요해보입니다. > | 식당 > 이전 > 중앙 | 불면 - 위생 | 2019-03-15 | 중앙식당에서 육계장을 먹으려고 기다리면 중에 육 | |
| 2019-03-14 오후 2:10 | 잠실-송파-북정 자율출퇴근 > | 통근/서플버스 > 이전 | 건의 | 2019-03-14 | 지금도 잠실에서 천로 상일동을 지나는 조기출근 버 | |
| 2019-03-13 오후 3:10 | 버스 탑승관련 문의 > | | | | 안녕하십니까! 항상 사원들의 편리한 교통수단을 위 | |
| 2019-03-13 오전 1:33 | 출근 버스 노선도 모바일 앱 > | 통근/서플버스 > 이전 | 불면 - 이상 | 운행 | 2019-03-13 | 3월 13일 수요일 신대방선 오전 06:30 통근 버스를 |
| 2019-03-13 오전 10:22 | happyroad 어플이름 관련 문 > | | | | 서플버스를 이용하고 있는데 이제 노선 및 시간 검 | |
| 2019-03-12 오후 3:59 | 분실물문의 > | | | | 19.03.12일 (금일) 수영장에서 소니 방수이어폰을 잃 | |
| 2019-03-12 오전 9:01 | 하남 미사 지역 조출버스 중 > | 통근/서플버스 > 이전 | 건의 | 2019-03-12 | 하남 미사지역 인원이 늘어 버스가 딱딱 타 | |
| 2019-03-11 오후 2:38 | 통근 버스 개선 요청드립니다 > | | | | 통근 버스 카테고리 이동하여 다시 올립니다.청주- | |
| 2019-03-09 오후 10:04 | 오픈마켓 > | 식당 > 청주 > 1 | 행퍼 불면 - 맛&품질 | 2019-03-08 | 매주 금요일날 나오면 오픈마켓은 1인 4기 무료로 바 | |
| 2019-03-09 오후 6:31 | 박실장님 칭찬합니다. > | 식당 > 이전 > 복지 | 칭찬 | 2019-03-09 | 맛있는 반찬에 대해 물어보니 친절하게 답변해 주 | |
| 2019-03-09 오후 6:27 | 박실장님 칭찬합니다. > | | | | 반찬에 대해 물어 보니 친절하게 답변주셔서 감사 | |
| 2019-03-09 오전 11:45 | 집에 면발로 갈 뻔 했어요 > | 통근/서플버스 > 이전 | 불면 - 기타 | 2019-03-08 | 우리 출퇴근 버스는 점심때만요요요요요요요요 | |

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|-----|-----|----|-----|-----|-----|------|------|-----|-----|-----|-----|-----|
| 서울 | 작석 | 고장 | 작곡 | 통로 | 죽 | 번재 | 음 | 좌석 | 고정 | 등반이 | 뒤 | 배자 |
| 통신사 | 통신사 | 사용 | 중 | 확인 | 통신망 | 설치 | 못하 | 좌석 | 고정 | 등반이 | 뒤 | 배자 |
| 히터 | 기사님 | 겨울 | 셔츠 | 발목 | 없 | 드검 | 발 | 히터 | 고정 | 등반이 | 뒤 | 배자 |
| 버스 | 탑승 | 관련 | 문 | 안녕 | 사원 | 편리 | 교통수단 | 노력 | 담당자 | 분 | 감사 | 출발 |
| 통근 | 버스 | 개선 | 요청 | 통근 | 버스 | 카테고리 | 이동 | 청주 | 통근 | 버스 | 이용 | 구성원 |
| 박 | 실장 | 칭찬 | 반찬 | 친절 | 답변 | 감사 | | | | | | |
| 박 | 실장 | 칭찬 | 반찬 | 친절 | 답변 | 감사 | | | | | | |
| 집 | 연발 | 변 | 근 | 버스 | 침침 | 방인 | 가요 | 요즘 | 날씨 | 출발 | 시간 | 발 |
| 통근 | 버스 | 환경 | 개선 | 부탁 | 청주 | 통근 | 수선 | 상황 | 정장 | 복지회 | 관 | 식당 |
| 복지 | 회관 | 식당 | 밥 | 맛 | 줄 | 박 | 실장 | 친절 | 정장 | 복지회 | 관 | 식당 |
| 박태준 | 실장 | 칭찬 | 아침 | 행 | 코너 | 변경 | 수선 | 상황 | 정장 | 복지회 | 관 | 식당 |
| 히터 | 중 | 히터 | 요즘 | 날씨 | 줄 | 가요 | 요즘 | 날씨 | 출발 | 시간 | 발 | 발 |
| 기사 | 아저씨 | 감사 | 오전 | 밥 | 출근 | 버스 | 도로 | 위 | 기사 | 아저씨 | 주선 | 운행 |
| 진리 | 정류장 | 미 | 청차 | 변 | 차 | 청 | 안 | 환수 | 말 | 출퇴근 | 시 | 발 |
| 버스 | 히터 | 클 | 선 | 배발 | 기사님 | 시간 | 급출발 | 급출발 | 못박 | 시내 | 레이상 | 얼미 |
| 저녁 | 답 | 선 | 배발 | 기사님 | 시간 | 급출발 | 급출발 | 못박 | 시내 | 레이상 | 얼미 | 상하 |
| 버스 | 대내 | 금지 | 해 | 주세 | 통역구 | 청 | 미 | 정차 | 포환 | 출 | 날 | 버스 |
| 청주 | 고대 | 조 | 출근 | 버스 | 통역구 | 청 | 미 | 정차 | 포환 | 출 | 날 | 버스 |
| 학교 | 선 | 버스 | 기사 | 안녕 | 하세 | 학교 | 선 | 버스 | 사원 | 중 | 한명 | 조 |
| 고대 | 통근 | 버스 | 청주 | 운임 | 통 | 관련 | 상기 | 건 | 관련 | 고객 | 소리 | 글 |
| 청주 | 고대 | 버스 | 노선 | 기사분 | 설날 | 아침 | 기분 | 나쁘 | 기분 | 조 | 퇴근 | 집 |
| 차 | 안 | 답 | 전 | 글 | 답 | 없 | 운전 | 기사님 | 셔츠 | 조끼 | 기모 | 티 |
| 서울 | 버스 | 관련 | 기사님 | 덕분 | 편하 | 서울 | 버스 | 이용 | 서울 | 버스 | 작성 | 전 |
| 서울 | 버스 | 히터 | 출 | 답 | 히터 | 출발 | 시 | 뒤 | 관련 | 꺼 | 날 | 버 |
| 차내 | 답 | 정도 | 서울 | 퇴근 | 두꺼운 | 점퍼 | 히터 | 세 | 조 | 절 | 부탁 | |
| 봉명 | 기숙사 | 술 | 세요 | 안녕 | 하세 | 봉명 | 기숙사 | 사람 | 다르 | 오늘 | 새벽 | 잠 |
| 수도 | 관광 | 버스 | 수도 | 관광 | 버스 | 노후 | 말미 | 작성 | 수도 | 부족 | 후 | 뒷 |
| 수도 | 관광 | 버스 | 수도 | 관광 | 버스 | 노후 | 말미 | 작성 | 수도 | 부족 | 후 | 뒷 |

- 대상 : SK 하이텍 고객의 소리 게시판
 - 기간 : 2008년 4월 4일 ~ 2019년 3월 21일
 - 개수 : 5,139개
- (총 9,240개 중 비밀글 4,101 개를 제외)

- 변수명
- 1) Datetime
 - 2) Title
 - 3) Text
 - 4) Voc_type
 - 5) Voc_date
 - 6) Service

- Parsing 대상 : Title & Text
- Parsing 방법 : 꼬꼬마 형태소 분석기
- 형태소 필터링 : 일반명사, 고유명사,

2. Topic Modeling

Parsed Data

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|-----|-----|-----|-----|-----|-----|------|------|-----|-----|-----|-----|-----|
| 서울 | 좌석 | 고장 | 좌측 | 통로 | 쪽 | 번째 | 쯤 | 좌석 | 고정 | 등받이 | 뒤 | 배차 |
| 통신사 | 사용 | 통신사 | 중 | 확인 | 통신망 | 설치 | 못하 | 사용자 | 공장 | 핸드폰 | 사용 | 금 |
| 히터 | 기사님 | 겨울 | 서즈 | 발목 | 없 | 뜨겁 | 발 | 히터 | | | | |
| 버스 | 탑승 | 관련 | 문 | 안녕 | 사원 | 편리 | 교통수단 | 노력 | 담당자 | 분 | 감사 | 출발 |
| 통근 | 버스 | 개선 | 요청 | 통근 | 버스 | 카테고리 | 이동 | 청주 | 통근 | 버스 | 이용 | 구성원 |
| 박 | 실장 | 칭찬 | 반찬 | 친절 | 답변 | 감사 | | | | | | |
| 박 | 실장 | 칭찬 | 반찬 | 친절 | 답변 | 감사 | | | | | | |
| 집 | 맨발 | 뻘 | 근 | 버스 | 찜질 | 방인 | 가요 | 요즘 | 날씨 | 출발 | 시간 | 발 |
| 통근 | 버스 | 환경 | 개선 | 부탁 | 청주 | 통근 | 버스 | 이용 | 구성원 | 최근 | 복대등 | 금호 |
| 복지 | 회관 | 식당 | 밥 | 맛 | 출 | 박 | 실장 | 친절 | 칭찬 | 복지회 | 관 | 식당 |
| 박태준 | 실장 | 칭찬 | 아침 | 행 | 코너 | 변경 | 수선 | 상황 | 무 | 어떨 | 선택 | 당황 |
| 히터 | 중 | 히터 | 요즘 | 날씨 | 을 | 가요 | 덥 | 저번 | 을 | 여 | 주선 | 기서 |
| 기사 | 아저씨 | 감사 | 오전 | 앞 | 출근 | 버스 | 도로 | 위 | 기사 | 아저씨 | 버스 | 운행 |
| 진리 | 정류장 | 미 | 정차 | 번 | 차 | 정 | 차 | 안 | 횃수 | 많 | 불편 | 뻘 |
| 버스 | 히터 | 글 | 번 | 출 | 글 | 본적 | 담당자 | 님 | 없 | 출퇴근 | 시 | 발 |
| 저녁 | 답 | 선 | 백발 | 기사님 | 시간 | 급출발 | | 뜻밖 | 시내 | 레이싱 | 멀미 | 심하 |
| 버스 | 매너 | 공지 | 해 | 주세 | 날씨 | 외투 | 두껍 | 서울 | 명이 | 자리 | 좁 | 서로 |
| 청주 | 교대 | 조 | 출근 | 버스 | 통역구 | 청 | 미 | 정차 | 포함 | 좁 | 날 | 버스 |
| 향교 | 선 | 버스 | 기사 | 안녕 | 하세 | 향교 | 선 | 버스 | 사원 | 중 | 한명 | 조 |
| 교대 | 통근 | 버스 | 청주 | 용암동 | 기사님 | 관련 | 상기 | 건 | 관련 | 고객 | 소리 | 글 |
| 청주 | 교대 | 버스 | 노선 | 기사분 | 실날 | 아침 | 기분 | 나쁜 | 기분 | 좋 | 퇴근 | 집 |
| 차 | 안 | 답 | 전 | 글 | 답 | 없 | 운전 | 기사님 | 서즈 | 조끼 | 기모 | 티 |
| 서울 | 버스 | 관련 | 기사님 | 덕분 | 편하 | 서울 | 버스 | 이용 | 서울 | 버스 | 착석 | 전 |
| 서울 | 버스 | 히터 | 출 | 덥 | 히터 | 출발 | 시 | 뒤 | 괜찮 | 꺼 | 기사님 | 전달 |
| 차내 | 덥 | 정도 | 서울 | 퇴근 | 두꺼운 | 정퍼 | 히터 | 세 | 조 | 절 | 부탁 | |
| 봉명 | 기숙사 | 술 | 세요 | 안녕 | 하세 | 봉명 | 기숙사 | 사람 | 다르 | 오늘 | 새벽 | 좁 |
| 수도 | 관광 | 버스 | 수도 | 관광 | 버스 | 노후 | 멀미 | 좌석 | 수도 | 부족 | 후륜 | 뒷 |
| 수도 | 관광 | 버스 | 수도 | 관광 | 버스 | 노후 | 화로 | 사내 | 서울 | 이용 | 시 | 멀미 |

LDA Topic Modeling

Topic 1 Data

Topic 2 Data

⋮

Topic K Data



2. Topic Modeling

Topic 2: 버스 노선, 운행 불만/건의

지금도 잠실에서 천호 상일동을 지나는 조기**출근** 버스가 만차되어 매일 10~20명 정도의 인원이 택시를 타고 **출근**하고 있다고 합니다. 잠실에서 송파를 지나서 복정까지 가는 자유훈**출퇴근** 조기**출근** 버스가 추가로 생긴다면 기존의 **노선** 만차 문제 ...

수지선 통상 **출근버스** 이용자입니다.고담 주차장으로 **셔틀**정차지가 바뀐 후 부터 **출근버스**의 도착**시간**이 평균 8:10 이후 입니다.특히 월요일의 경우엔 8:20다되어 도착하기도 합니다 ...

잠실선 교대 야간 **출근버스** 매번 탈때마다 택시타고 **출근**합니다. 사람도 많이 이용하는데 **증차**시켜주세요.택시도 3명씩 타는것도 아니고 4명씩 꽉 채워서 가고 **불편**합니다 ...



2. Topic Modeling

Topic 3: 버스, 기숙사, 식당 문의

이번 M16 Set-up 관련 M10F 건물 앞 **사내 셔틀** 정류장이 철거되어 현재 **셔틀**을 타려면, 기술지원센터 또는 경영지원 본부 앞까지 가서 타야하는 상황입니다. **교대 셔틀**은 M10F 사거리 횡단 보도 앞에서 하차하는데, 왜 **사내 셔틀**은 이용하지 못하는지 ...

안녕하세요 사원 **아파트** 관련해서 문의 드립니다.. 사원 **아파트** 관리비 고지서를 받았는데 난방비가 11만원 이상 나왔습니다. 중앙 난방이라 그러려니 합니다. **아파트**가 오래되서 추워서 ...

... 현재 결제법에 있어 이용자가 불편함과 불합리함을 감수할 부분이 아닙니다. 1. 페이코 **결제** 자체를 문제 삼는 것이 아니라 페이코 내부에도 여러가지 **결제** 방법이 있는데, 사용자 편의주의 ...



2. Topic Modeling

Topic 4: 식당 메뉴, 편의식 평가

편의식 건의드립니다.건강식도 좋지만 우선 배가 안차요.. 우선 배가 든든해야 일을 하는데 샌드위치 한두조각 먹고 배 안참니다. **빵** 두개 음료 두개씩 이라도 선택할수 있게 건의드립니다.**빵**하나 음료하나 그리고 바나나+아무거나... 이젠 질립니다 ...

금일 석식을 먹으러 갔는데**메뉴** 변동이 왜이렇게 심한가요?새우 튀김과 오븐콘치즈가 나온다고 해서 간건데해당 메인디쉬는 다 빠져있고 대체 **음식**으로 굴 하나, 요구르트 하나 주네요 ...

안녕하세요.**중앙식당**을 자주 이용하고있는 구성원입니다.4끼 모두 편의식을 받을 수 있게 되어 좋은데요.현재 4층 진열대에서 층별로 1개씩 가져가야 하는데맨위 **빵**(밀가루..) 외에도 ...



2. Topic Modeling

Topic 5: 구성원에 대한 평가

김반장님 및 기사님들 감사합니다. 통근버스 주차장 이동 후 사내 순환셔틀 편리하게 이용하고 있습니다. 항상 친절하게 운행하시는 모습에 늘 감사하는 마음은 있었으나 표현하지 못하고 있었습니다. 이 자리를 빌어 감사의 마음 전합니다.

오늘 오후 5시반에서 5시50분 정도 사이에 M15 정문 가운데 옆에서 검색하시던 안경쓰고 계셨던분 제가 정말 보안검색을 받고 있어도 너무나 감사하게 감사합니다. 수고하셨습니다...

청주 1캠 보안요원 여사원 관련 하여 말씀드립니다. 정문으로 출근 시 왼쪽에 앉아 계시는 여사원분, (타각기쪽에 더 가깝게 앉으신분.) 굉장히 불친절한데, 컨셉인가요??? 버스에서 ...



3. Sentiment Analysis

1. 파싱 데이터(형태소 매칭)

| 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|--------|--------|--------|--------|--------|--------|--------|
| 서툰/NNG | 좌석/NNG | 고장/NNG | 좌측/NNG | 통로/NNG | 쪽/NNB | 번패/NNB |
| 청운/NNG | 중/NNG | 불편/NNG | 점/NNG | 많/VA | 설날/NNG | 전/NNG |
| 편/NNG | 의식/NNG | 선택/NNG | 편/NNG | 의식/NNG | 건의/NNG | 건강식/NN |
| 통신사/NN | 통신사/NN | 사용/NNG | 중/NNB | 확인/NNG | 통신량/NN | 설치/NNG |
| 기분/NNG | 요/NNG | 아침/NNG | 기분/NNG | 좋/VA | 테이크/NN | 아웃/NNG |
| 건강식/NN | 샐러드/NN | 품/NNG | 소스/NNG | 관련/NNG | 기타/NNG | 의견/NNG |
| 다이어트/1 | 메뉴/NNG | 부활/NNG | 요청/NNG | 건/NNG | 다이어트/1 | 메뉴/NNG |
| 히터/NNG | 기사님/NN | 겨울/NNG | 셔츠/NNG | 발목/NNG | 없/VA | 뜨겁/VA |
| 식단/NNG | 일본어/NN | 사용/NNG | 필요/NNG | 가요/NNG | 오/NNG | 라이스/NN |
| 행복/NNG | 마을/NNG | 행복/NNG | 마을/NNG | 맛있/VA | 것/NNB | 같/VA |
| 요플레/NN | 오늘/NNG | 이번/NNG | 주/NNG | 요플레/NN | 출/NNB | 요플레/NN |
| 동/NNG | 퇴근/NNG | 버스/NNG | 노선/NNG | 숙지/NNG | 출발/NNG | 동/NNG |

2. 감성사전

Home

Corpus

Lexicon

Publications

Korean Sentiment Analysis Corpus

About

The Korean Sentiment Analysis Corpus, KOSAC, is built for capturing sentiment expressions and their patterns in Korean and representing their meaning to be interpretable for a computer. To represent the meaning, a fine-grained annotation scheme called KSMAL (Shin et al., 2012) is developed identifying key components and properties of sentiments based on solid theoretical background. The annotation scheme has been employed in the manual annotation of a 7,713-sentence corpus of 332 news articles from the Sejong syntactic parsed corpus.

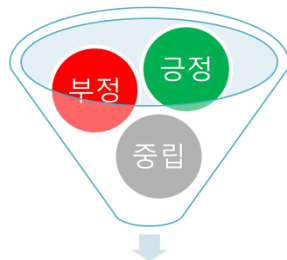
Annotated Sentences

| Subjective | Objective | Total |
|------------|-----------|-------|
| 2655 | 5055 | 7713 |

Annotated Seed Tags

| | Agree | Argo | Emotion | Intention | Judgment | Speculation | Others | Total |
|----------------|-------|------|---------|-----------|----------|-------------|--------|-------|
| Dir-Action | 1 | 9 | 71 | 8 | 38 | 0 | 1 | 128 |
| Dir-Explicit | 155 | 277 | 341 | 276 | 2740 | 157 | 40 | 3987 |
| Dir-Speech | 8 | 1149 | 22 | 28 | 96 | 13 | 7 | 1313 |
| Indirect | 255 | 321 | 720 | 409 | 6085 | 63 | 22 | 7676 |
| Writing-Device | 5 | 58 | 9 | 396 | 754 | 172 | 2657 | 4311 |
| Total | 425 | 1854 | 1163 | 1827 | 9714 | 405 | 3027 | 17618 |

3. 감성사전과 매칭



극성 추출

4. 게시글의 감성 예측

| index | neg | pos | sent_diff | tot | sent | sent_total |
|-------|-----|-----|-----------|-----|-----------|------------|
| 0 | 4 | 4 | 0 | 8 | 0 | 중립 |
| 1 | 5 | 6 | 1 | 11 | 0.090909 | 양성 |
| 2 | 5 | 4 | -1 | 9 | -0.111111 | 부정 |
| 3 | 6 | 4 | -2 | 10 | -0.2 | 부정 |
| 4 | 9 | 15 | 6 | 24 | 0.25 | 양성 |
| 5 | 8 | 9 | 1 | 17 | 0.058824 | 양성 |
| 6 | 2 | 2 | 0 | 4 | 0 | 중립 |
| 7 | 1 | 3 | 2 | 4 | 0.5 | 양성 |
| 8 | 5 | 3 | -2 | 8 | -0.25 | 부정 |
| 9 | 2 | 5 | 3 | 7 | 0.428571 | 양성 |
| 10 | 2 | 1 | -1 | 3 | -0.333333 | 부정 |
| 11 | 5 | 3 | -2 | 8 | -0.25 | 부정 |
| 12 | 5 | 8 | 3 | 13 | 0.230769 | 양성 |

3. Sentiment Analysis

1. 데이터 입력

아침 빵 메뉴가 **개선** 되서 너무 **좋**아요~ 안 그래도 며칠 전에 양을 줄이고 **질**을 높였으면 **좋**겠다고 생각했었는데 **즐겁**게 아침 식사 했습니다^
^ 그런데 프리미엄 우유도 **조금** 다양화 하면 **좋**지 않을까요? 예를 들면 같은 회사 제품인 드링킹 요구르트나.. 아무튼 수고가 많으세요!^^

2. 파싱

아침/빵/메뉴/개선/좋/며칠/전/양/질/좋/생각/즐겁/아침/식사/프리미엄/우유/조금/다양화/하면/좋/예/같/회사/제품/링/킹/요구르
트/수고/많/

3. 긍정/부정 분류

긍정 : 개선, 좋, 질, 좋, 즐겁, 좋

부정 : 조금

4. 감성분석 예측

약 0.85의 극성값으로 위의 게시물을 '긍정'으로 판단.

4. Web Interface

Topic 1 Data

Topic 2 Data



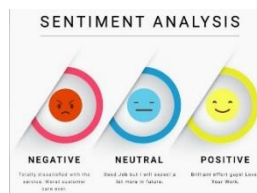
Topic K Data



Trend Chart : 시간에 따른 검색량의 변화를 확인



Top Keywords : 검색량 상위 키워드 추출 및 워드클라우드

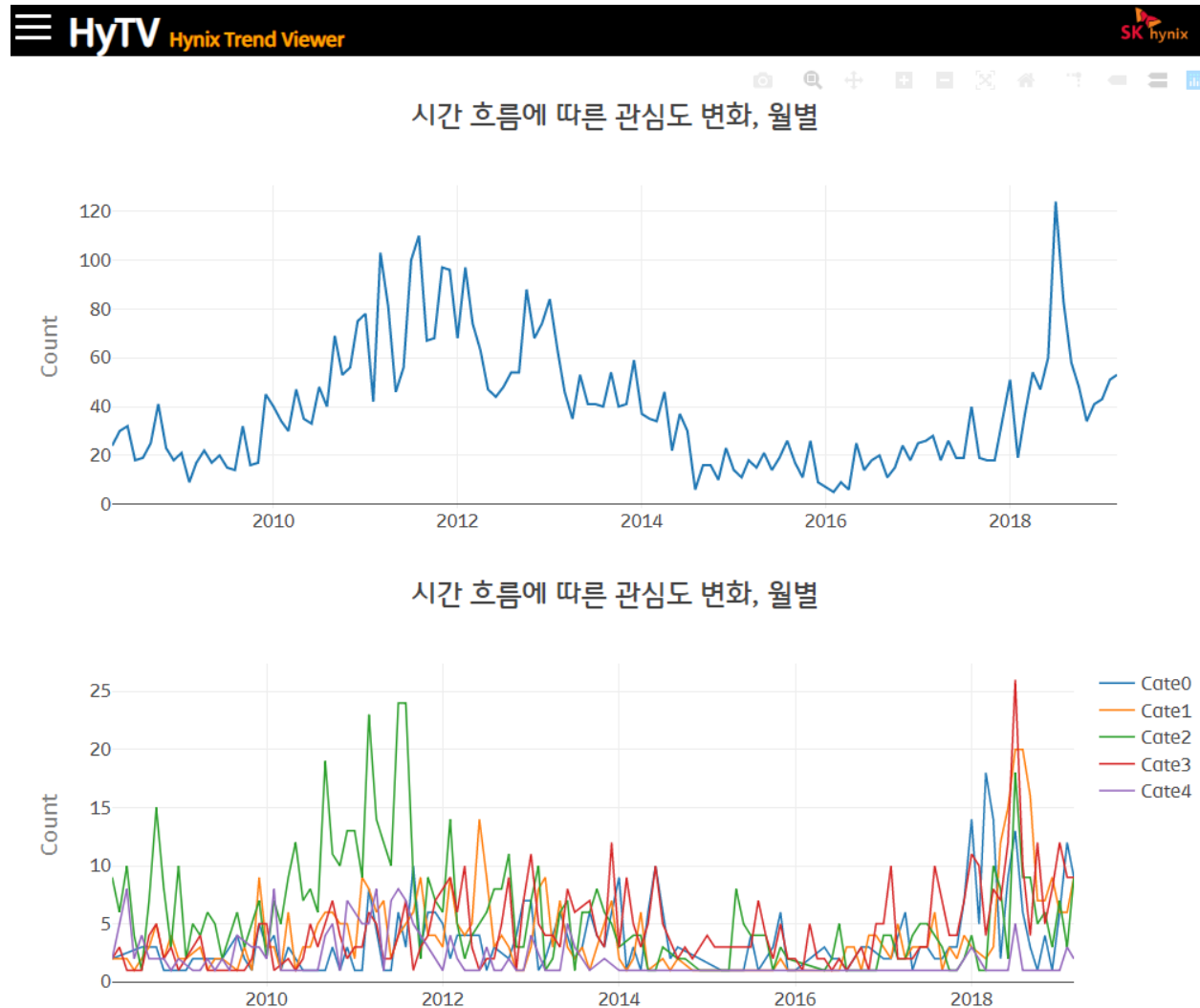


Sentiment Analysis : 토픽에 대한 구성원의 인식을 확인

4. Web Interface

Trend
Chart

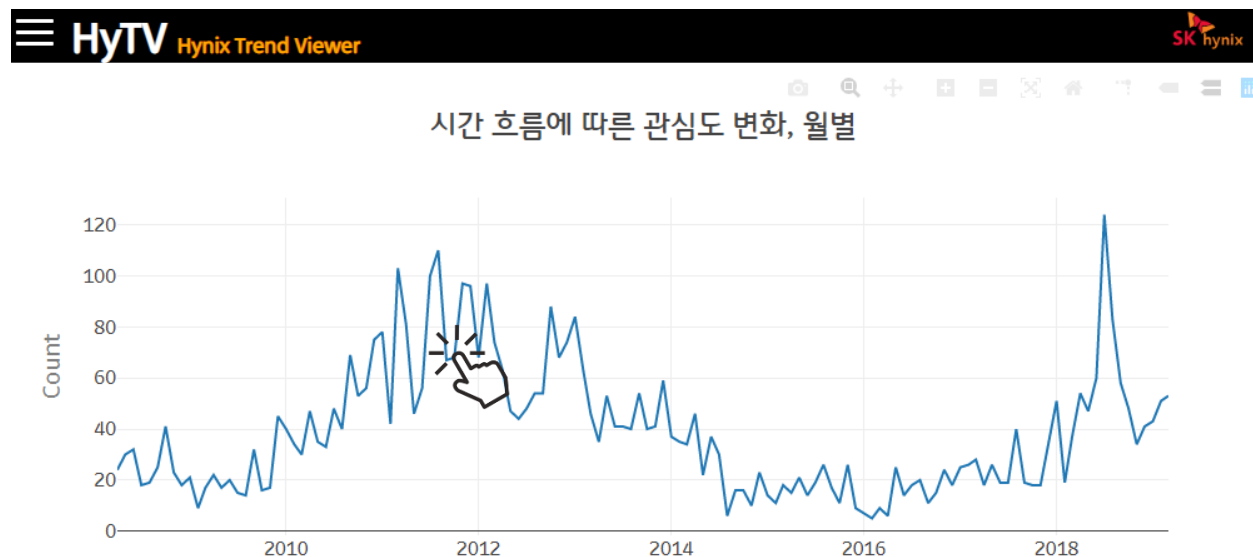
검색 빈도 수 차트



4. Web Interface

Trend
Chart

특정 시점의 상위 키워드 확인

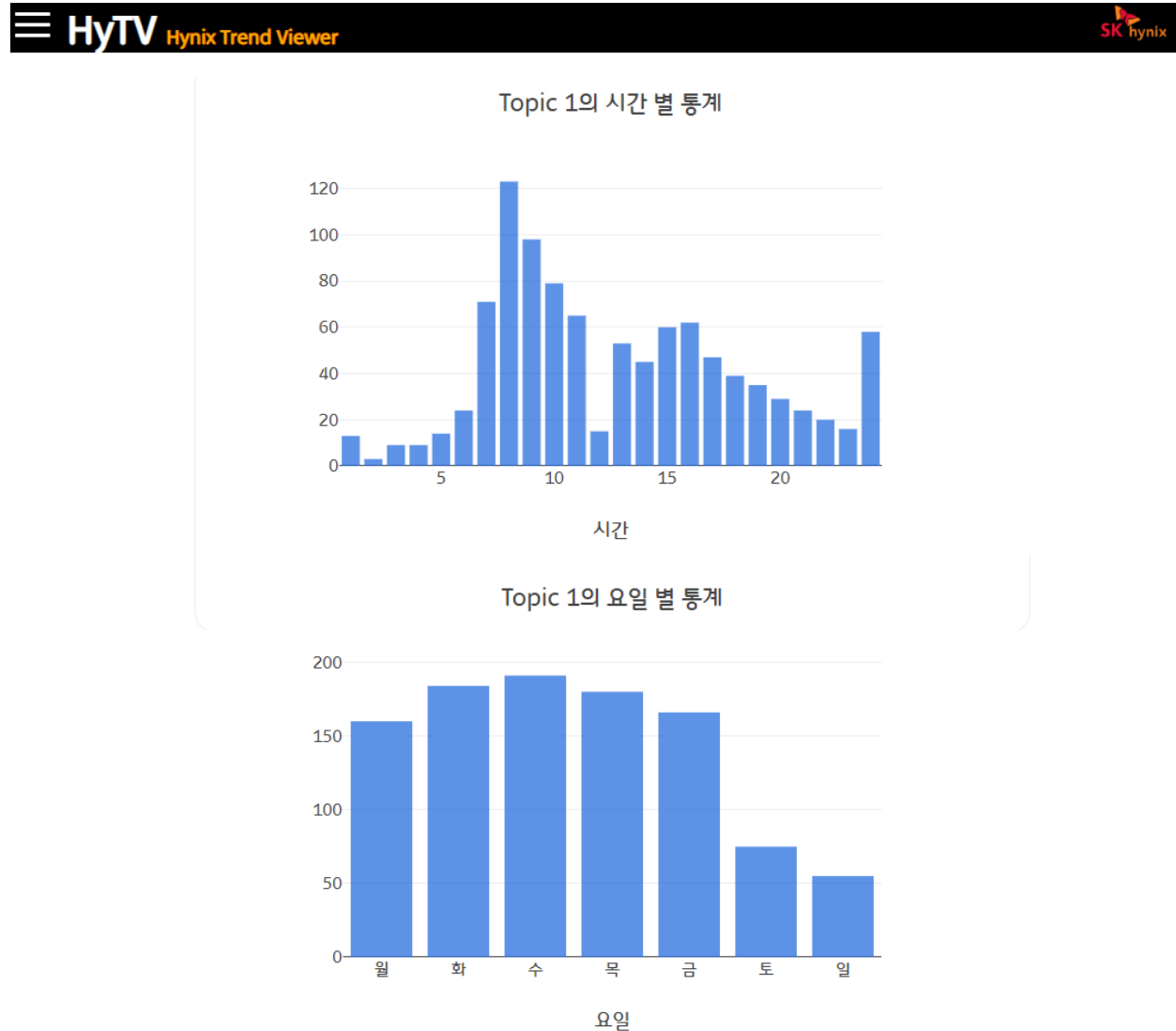


| Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 |
|-----------|----------|----------|----------|----------|
| 1: 버스 28 | 1: 버스 48 | 1: 아파트 9 | 1: 식당 32 | 1: 분 14 |
| 2: 출발 23 | 2: 노선 26 | 2: 롯데 5 | 2: 밥 14 | 2: 버스 11 |
| 3: 시간 17 | 3: 퇴근 21 | 3: 낙천 5 | 3: 청운 14 | 3: 좌석 11 |
| 4: 출근 14 | 4: 인천 20 | 4: 대 5 | 4: 맛 12 | 4: 예약 10 |
| 5: 기사님 10 | 5: 선 18 | 5: 벽산 4 | 5: 빵 12 | 5: 농약 10 |

4. Web Interface

Trend Chart

시간, 요일 별 검색 빈도 확인



4. Web Interface

Top
Keywords

토픽 별 키워드 Top 5

| HyTV Hynix Trend Viewer | | SK hynix |
|-------------------------|-------|---|
| 키워드 Top 5 | | |
| 1: 버스 | 2921회 |  |
| 2: 기사님 | 992회 | |
| 3: 출근 | 780회 | |
| 4: 분 | 636회 | |
| 5: 통근 | 547회 | |

4. Web Interface

워드클라우드

Top
Keywords



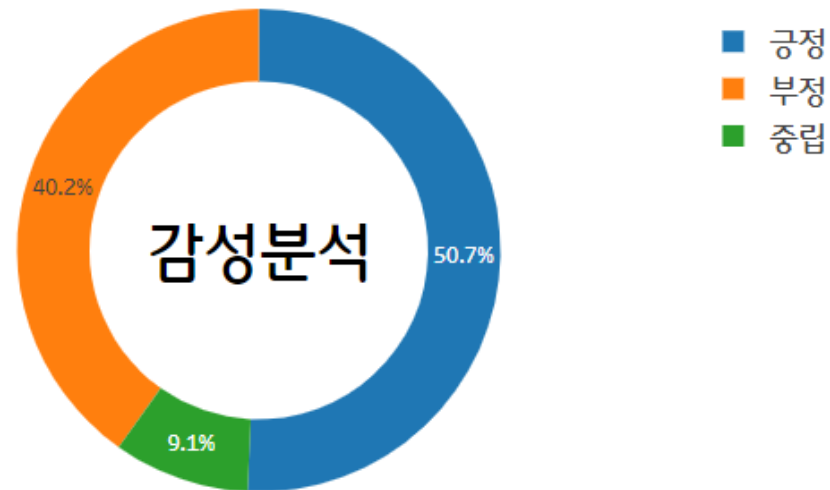
4. Web Interface

Sentiment
Analysis

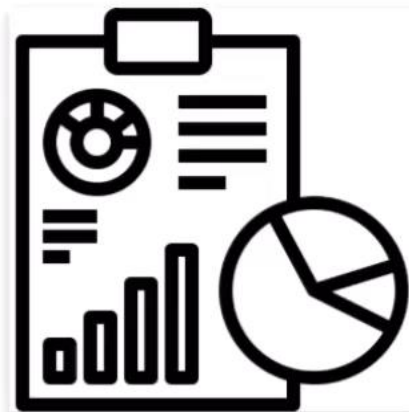
토픽 별 키워드 Top 5



Topic 1



Demonstration



Summary

- 프로젝트의 목적

- ✓ 하이스텍 고객의 소리 게시판의 글을 분석하여
- ✓ 시각화 결과를 Web Interface로 구축

- 프로젝트의 진행

1. Data Crawling
2. Data Preprocessing
3. Topic Modeling and Sentiment Analysis
4. Visualize
5. Web Interface 구축

- 프로젝트의 의의

- ✓ 관리자 입장 : 고객의 Needs를 빠르고 쉽게 확인할 수 있음
- ✓ 사용자 입장 : 다른 고객들의 관심사, 인식을 쉽게 파악할 수 있음

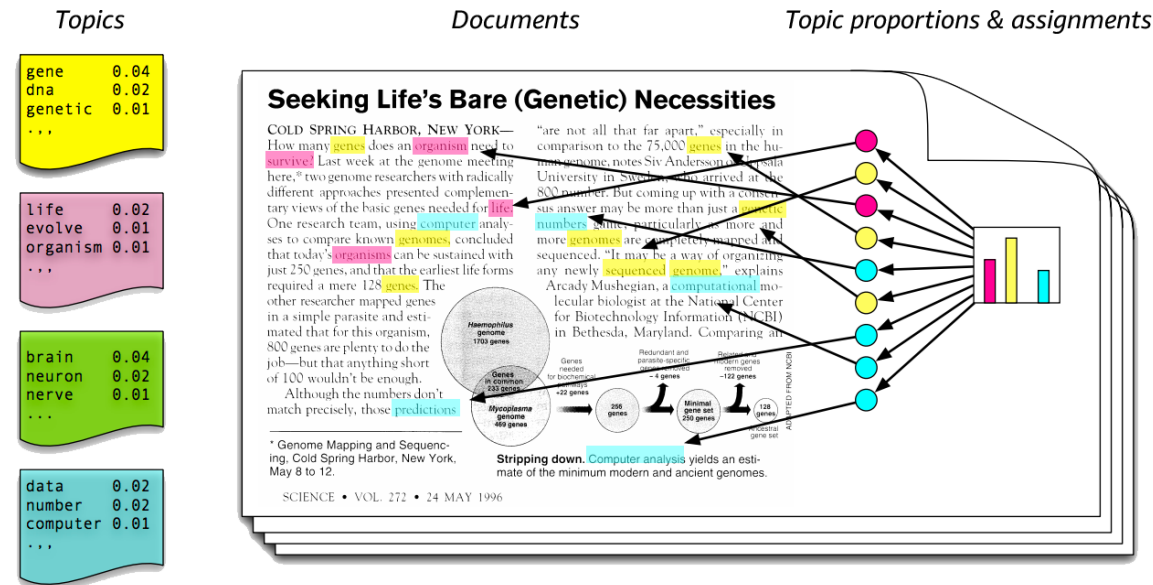
- Future Works

- ✓ Crawling의 자동화
- ✓ HyTV를 하이닉스 사내 게시판 (하이통)에 적용

Thank you

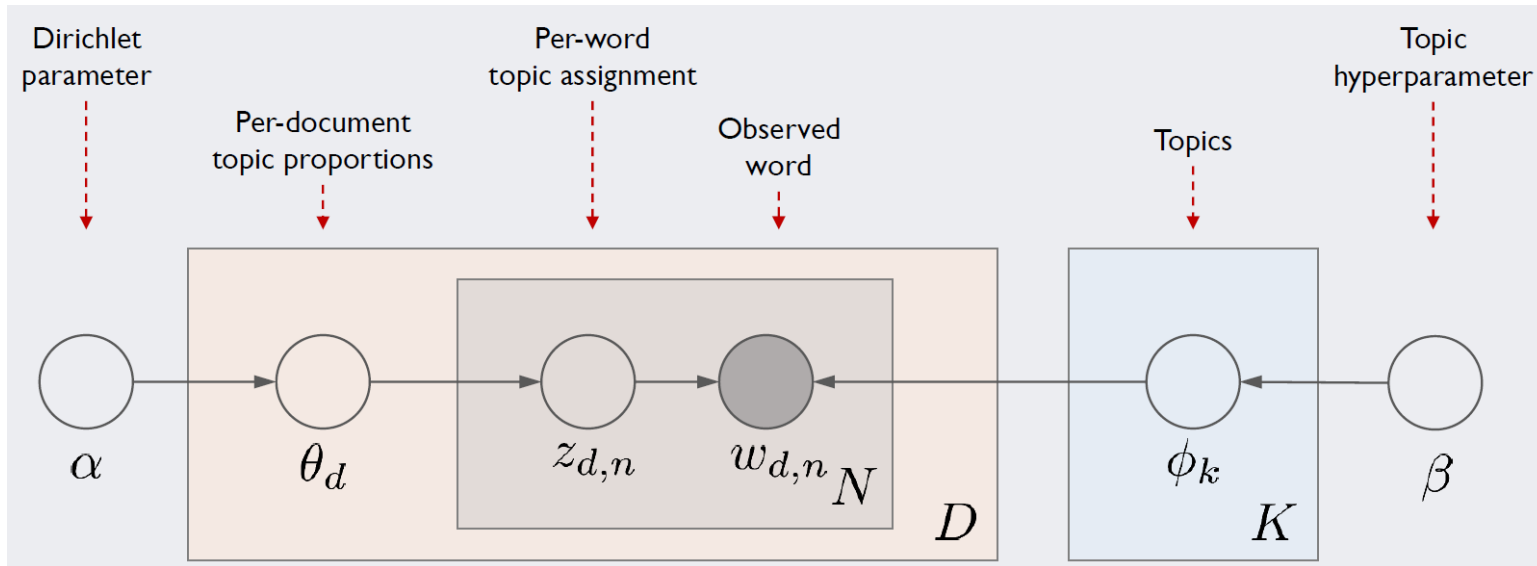
Latent Dirichlet Allocation (LDA)

- 주어진 문서에 대하여 각 문서가 가질 수 있는 토픽(주제)들에 대한 확률모형
- 문서별 토픽의 분포, 토픽별 단어의 분포를 추정하여 각 문서에서 우세한 토픽을 확인할 수 있음



Latent Dirichlet Allocation (LDA)

- LDA가 가정하는 문서 생성 과정



- D : 말뭉치 전체 문서 개수
- K : 전체 토픽 수
- N : d 번째 문서의 단어 수
- $w_{d,n}$: d 번째 문서에 등장한 n 번째 단어

Sentiment Analysis

- 감성사전으로 감성점수를 산출하는 방법

- 서울대 컴퓨터 언어학과에서 만든 KOSAC 기반의 Korean Sentiment Lexicon 감성사전을 이용 (<http://word.snu.ac.kr/kosac/>)
- Sentiment Indicators

$$polarity = \frac{p-n}{p+n}$$

$$subjectivity = \frac{n+p}{N}$$

$$pos\ refs\ per\ ref = \frac{p}{N}$$

$$neg\ refs\ per\ ref = \frac{n}{N}$$

$$senti\ diffs\ per\ ref = \frac{p-n}{N}$$

- p : 긍정 판단 단어 개수
- n : 부정 판단 단어 개수
- N : 전체 단어 개수