

## 프로젝트 최종 보고서

나는 숭실대학교 컴퓨터학부의 일원으로 명예를 지키면서 생활하고 있으며, 보고서를 작성하면서 다음과 같은 사항을 준수하였음을 엄숙히 서약합니다.

1. 나는 자력으로 보고서를 작성하였습니다.
2. 나는 보고서에서 참조한 문헌의 출처를 밝혔으며 표절하지 않았습니다.
3. 나는 보고서의 내용을 조작하거나 날조하지 않았습니다.

교과목	전공종합설계 II
교과목 담당교수	이 수 원

프로젝트명	SNS 텍스트데이터를 이용한 평판 분석
프로젝트 수행기간	2012. 9 ~ 2012. 12
프로젝트 지도교수	이 수 원
멘토 / 소속	한영우 / 한국예탁결제원
팀 명	blabla
팀 원	이봉기, 차소윤, 이호, 위석철
제출일	2012. 12. 18

## 목 차

1. 프로젝트의 개요 .....	01
2. 관련 기술 현황 .....	02
3. 프로젝트 설계 내용 .....	06
4. 최종 결과 .....	27
5. 평가 항목 및 방법 .....	38
6. 기대 효과 및 활용 방안 .....	43
7. 프로젝트 수행 체계 .....	44
8. 수상 현황 .....	50
9. 참고 문헌 .....	50
첨부 A. 회의록 .....	51
첨부 B. 소스코드 .....	91
첨부 C. 감성단어사전 .....	186

# 1. 프로젝트의 개요

## 1-1. 프로젝트 개발 배경

오늘날 많은 스마트폰이 개발 및 보급되고 있고, 이는 주로 개인용 컴퓨터의 웹 브라우저를 이용했던 과거의 인터넷 이용 방식에 큰 변화를 가져왔다. 얼마 전까지 언제 어디서나 휴대전화를 사용해 SMS를 이용했던 것처럼, 이제는 언제 어디서나 스마트폰을 사용해 SNS(Social Network Service)에 글을 남기는 것이 가능해졌다. 이에 따라 SNS는 급속도로 성장하게 되었고, 사용자 또한 급격하게 증가하였다. 지금도 수많은 사람들이 SNS에서 글을 쓰고 있고, SNS상에는 매일 엄청난 분량의 텍스트 데이터가 생성되고 있다.

이런 많은 양의 텍스트 데이터 안에는 개인의 사사로운 기분이나 일상부터 다른 사람들에게 알려주고 싶은 내용, 유용한 정보, 개인적인 생각에 대한 표출, 한 주제에 대한 여러 사람의 토론 등의 다양한 내용이 들어있다. 이런 다양한 내용을 살펴보면 영화나 공연 등에 대한 리뷰, 여러 브랜드에 대한 선호도, 관광지에 대한 만족도, 전자 제품 사용 후기 등 어떤 특정한 대상에 대한 의견들이 들어있다는 것을 알 수 있다.

특정한 대상에 대한 개인의 의견이 모이면 그 대상에 대한 여론 또는 평판이 형성되는데, 이런 평판의 내용을 형체가 없는 입소문이 아닌 읽을 수 있는 구체적인 형태로 나타내고 수치화하거나 도식화시켜 알아보기 좋게 만드는 것은 매우 의미 있는 일이다. 평판을 구체적이고 알아보기 좋은 결과물로 가공하면 개인은 특정 제품에 대한 평판을 보고 그 제품을 살지 말지 결정하거나 특정 단체에 가입할지 말지를 결정할 수 있는 등의 방식으로 사용할 수 있으며, 기업은 어떤 제품이 대중에게 선호되고 있는지를 조사해 그에 따라 어떻게 제품을 개발하고 마케팅 할지를 결정할 수 있고, 이미 개발한 제품에 대한 평판을 조사하여 개선하는 등의 일에 사용할 수 있다.

특정 대상에 대한 평판을 알기 위해서는 일반적으로는 직접 사람들에게 물어보거나 설문조사를 하는 등의 방법을 사용한다. 하지만 시대의 흐름에 따라 개인적인 의견을 표출하는 방법은 다양해지고 있고, 이런 의견을 모아 분석하기 위해서는 각각의 의견 표출 방법에 대한 가장 효율적인 분석 방법을 사용하는 것이 좋다. SNS에 표출되어있는 특정 대상에 대한 의견들을 수작업으로 분석하면 정확한 결과를 얻을 수는 있겠지만, 매우 많은 노력과 시간과 비용이 소요되고 다양한 정보를 분석하는 것도 쉽지 않으며 인위적으로 분석되기 때문에 분석 대상과 분석 결과에 대한 객관성이 떨어질 우려가 있다. 따라서 자동화 되어있어 수작업을 많이 거치지 않아도 되고, 특정 개인 또는 집단의 사고방식이나 업무방식 등에 영향을 받지 않고 객관적으로 평판을 분석하는 시스템이 필요해진다. 본 프로젝트(blabla)는 이런 시스템을 개발하는 것을 목표로 하며, 프로젝트 개발 기간과 난이도를 고려해 모든 대상에 대한 평판이 아닌 영화에 대한 평판만을 조사하는 것으로 하고, 특정 영화가 좋은 쪽으로 평가되고 있는지 나쁜 쪽

으로 평가되고 있는지를 조사해 그 영화에 대한 선호도를 평가하는 시스템을 구축하는 것으로 목표를 한정한다. 본 프로젝트에서는 분석 대상을 영화에 대한 평판으로만 한정했지만, 이와 같은 방식으로 다른 분야에 대한 평판분석도 가능하다.

## 1-2. 프로젝트 최종 목표

SNS상에 존재하는 많은 양의 텍스트 데이터에서 최근에 상영된 영화에 대한 의견을 수집해 각각의 영화와 영화가 가진 속성에 대한 선호도를 분석하고 분석된 정보를 웹페이지와 안드로이드 애플리케이션을 통해 개인 또는 기업에게 제공하는 영화 평판 분석 시스템을 구현한다.

## 1-3. 중심어

한 글	평판 분석	감정 분석	텍스트	소셜 네트워크 서비스	선호도
영 문	Opinion mining	Sentiment analysis	Text	Social network service	Preference

# 2. 관련 기술 현황

## 2-1. 국내외 기술 현황

### ○ 상품 리뷰의 속성별 긍정 / 부정 분류 시스템의 설계 및 구현 [1]

쇼핑몰의 상품평을 이용해 상품과 상품의 속성에 대한 선호도를 분석하는 방법에 관한 논문이다. 이전에 저장된 상품평 패턴과 일치하는 정도를 확인해 긍정적인 평인지 부정적인 평인지를 분석한다. 이 논문에서는 이전에 저장된 상품평 패턴과의 비교를 통해 선호도를 분석했지만, 본 프로젝트에서는 문장이 포함된 단어가 긍정적인 단어인지 부정적인 단어인지를 분석해 문장의 선호도를 분석한다.

### ○ 감성 단어 분석 모듈을 이용한 모바일 소셜 네트워크 서비스에서의 성향 판단 시스템 구현 [2]

모바일 환경에서 SNS상에 올라온 글이 긍정적인지 부정적인지 판단하는 시스템에 대한 논문이다. 기존에 만들어져있던 감정 단어 사전을 활용하고 감정 단어의 빈도수에 의해 감정에 가중치를 준 다음 글의 성향을 판단한다. 이 논문에서 제안된 시스템은 모바일 시스템에서 실행되고 따라서 시스템의 성능이 모바일 장치의 성능에 영향을 받지만, 본 프로젝트는 분석 작업이 서버에서 이루어지기 때문에 분석 결과를 보는 컴퓨터의 성능은 크게 문제되지 않는다. 그리고 이 논문에서 제안된 시스템은 글 자체의 성향을 판단하지만, 본 프로젝트는 특정한 대상에 대한 선호도를 분석한다.

### ○ 상품 리뷰 요약에서의 문맥 정보를 이용한 의견 분류 방법 [3]

상품리뷰를 대상으로 평판 분석을 수행하는 경우 어휘의 성향을 파악할 때 문맥정보를 활용하는데 이를 위해 어휘가 사용될 때의 문맥정보를 정의하고 이를 의견분류에 적용하는 방법을 제안하였으며, 실험을 통하여 기존 연구보다 상황별로 알맞은 의견분류가 가능함을 보였다. 또한 수작업으로 말뭉치의 핵심 어휘들을 정의했던 기존 연구들에서의 방식에서 벗어나, 리뷰 본문과 리뷰 점수를 활용하여 자동으로 상황에 맞는 말뭉치를 구축한다. 이를 통해 상품리뷰에 나타난 어휘들의 문맥에 맞는 의미 성향을 판별한다.

### ○ 형식적 및 비형식적 어휘 정보를 반영한 문장 감정 분류 [4]

단어의 사용 범위에 따라 단어의 의미가 긍정이 될 수도 있고 부정이 될 수도 있기 때문에 이 논문에서는 사용 범위별 감정 사전을 구축했다. 그리고 이모티콘 처리를 위해서 대표 이모티콘을 정하고 변형 이모티콘과 대표 이모티콘의 유사도를 분석해 가장 비슷한 이모티콘을 이용한다. 이후 문장의 특징벡터를 계산하고 SVM(Support Vector Machine) 분류 방법으로 분류한다.

### ○ 관객의 관람평 분석을 통한 영화의 특성 추출 [5]

영화의 관람평을 분석해 영화가 어떤 특징을 가지고 있는지를 알아내는 방법에 대한 연구이다. 영화의 속성에 대한 선호도를 나타내는 특정한 단어가 전체 관람평에서 사용된 빈도 비교해서 분석하려는 관람평에서 얼마나 자주 사용되었는지를 측정한다. 그리고 특정한 단어와 영화의 속성간의 전체 관람평에서의 상호 정보량과 분석하려는 관람평에서의 상호정보량의 비율을 측정한다.

## 2-2. 국내외 경쟁기관 현황

### ○ TrendUp

TrendUp은 타파크로스(<http://www.tapacross.co.kr>)에서 제공하는 SNS 분석 서비스로서 고도화된 언어처리 기술을 바탕으로 주제별 이슈와 감성을 추출하고, 소셜 채널과 인플루언서<sup>1)</sup>를 분석하여 브랜드와 관련된 이슈, 인식, 의견, 행동패턴 등을 도출함으로써 마케팅, 홍보, CS 전략 활용을 목적으로 하고 있다. 지난 3월, 교보문고에서는 소비자와의 실시간 소통을 통해 고객 만족도 및 서비스 품질 향상을 위해 트렌드업을 도입하였다.

### ○ Pulse-K

Pulse-K는 코난테크놀로지(<http://www.konantech.co.kr>)에서 출시한 평판 분석 서비스로서 자연어 처리 및 텍스트 마이닝 기술로 개발한 웹 기반의 소셜 미디어 분석 서비스이다. SNS와 블로그 등에서 기업, 브랜드 그리고 제품에 대한 소비자들의 의견을 감성 분석하여 긍정·부정 의견으로 구분하여 모니터링을 할 수 있으며, 그 외에 실시간 이슈 키워드 출력이나 온라인 상에서 영향력 지수가

1) 인플루언서(Influencer) : SNS상에서 영향력이 있는 사용자

높은 인플루언서 발굴 등의 기능도 제공한다. 주 대상은 기업이며, 현재 오픈 베타 테스트를 진행하고 있다.

## ○ Smart Cruncher

ACTSMO(<http://www.actsmo.co.kr>)의 Smart Cruncher는 소셜 빅 데이터 분석 기술을 바탕으로 트위터, 블로그, 카페, 다양한 사이트 등 소셜 공간을 정성적/정량적으로 분석하여 customer insight를 확인할 수 있는 서비스이다. 키워드에 대한 사용자의 긍·부정적인 의견 확인과 Forecast, Big Mouth, 트렌드, 선호도, SWOT 분석<sup>2)</sup> 등의 다양한 기능을 제공한다. 소셜 데이터 감성분석 서비스는 지난 올레 벤처 어워드 2011 클라우드 서비스 분야 우수상 수상의 실적을 가지고 있다.

## ○ Adcoop

Adcoop는 유엑스코리아(<http://www.uxkorea.com>)에서 출시한 SNS 프로그램이다. 프로그램 내 기능으로 존재하는 SNS 분석 평판관리는 트위터상의 트윗을 수집하여 키워드별로 노출 빈도 수를 추출하고, 긍·부정 트윗을 분석하여 자료로 활용할 수 있다. 또 분석한 키워드에 대한 영향력이 큰 트위터 유저의 목록을 제공하는 기능을 가지고 있다. 주 대상은 기업이지만, 파워블로거들의 블로그 운영에도 많이 사용되고 있다.

## ○ BEOLTONG

BEOLTONG(<http://beoltong.wo.tc>)은 창원대에서 개발한 시스템으로, 트위터의 텍스트 데이터를 분석해 특정한 대상과 그 대상의 속성에 대한 선호도를 분석한다. 예를 들면 아이폰을 검색하면 아이폰 자체에 대한 선호도와 아이폰의 카메라, 디자인 등에 대한 선호도가 출력되고 분석 대상이 된 트위터의 글들이 출력된다. 현재 본 프로젝트의 기능과 가장 유사한 기능을 제공하지만, 속성에 대한 기준이 명확하지 않고, 미리 지정된 키워드에 대해서만 원활한 분석이 가능한 단점이 있다.

## ○ Tweet Sentiments

Tweet Sentiments(<http://tweetsentiments.com>)는 트위터의 텍스트 데이터를 분석해 입력된 키워드에 대한 선호도를 분석한다. 전 세계의 데이터를 분석하고, 최근의 데이터만 분석한다.

## 2-3. 국내외 지식재산권 현황

### ○ 상품평 극성 분류를 위한 특징별 서술어 긍정 및 부정 사전 자동 구축 시스템 및 방법 [6]

상품평의 긍정 또는 부정 서술어를 추출하고, 추출된 서술어와 단어의 접속정보와 평점을 이용해 단어의 긍정 또는 부정 여부를 분석한다. 본 프로젝트에서는 수작업으로 긍정 및 부정 사전을 직접 구축하지만, 차후 좀 더 많은 단어가 필요하게 되면 긍정 및 부정 사전을 자동으로 구축해 그 데이터를 활용하도록 할 가능성도 있다.

---

2) SWOT 분석 : 기업의 환경분석을 통해 강점(Strength)과 약점(Weakness), 기회(Opportunity), 위협(Threat) 요인을 규정하고 이를 토대로 마케팅 전략을 수립하는 기법

## 2-4. 기존 제품과의 차별성

구 분	선호도	키워드와 관계된 의견 수	도표 · 그래프	선호도 추이	보고서 생성	대상의 속성에 대한 선호도	대상 제한
TrendUp	○	○	○	○	○	×	제한 없음
Pulse-K	○	○	○	○	○	×	제한 없음
Smart Cruncher	○	○	○	○	○	×	제한 없음
Adcoop	○	○	○	○	×	×	제한 없음
BEOLTONG	○	○	○	○	×	○	제한 없음
Tweet Sentiments	○	○	○	×	×	×	제한 없음
Blabla	○	○	○	×	×	○	영화

[표 1] SNS 평판 분석 서비스 특징 비교

[표 1]의 SNS 평판 분석 서비스들을 살펴보면 모든 서비스가 입력한 키워드에 대한 선호도와 키워드와 관계된 의견의 수를 나타내고 도표와 그래프를 생성한다. 일부 서비스는 선호도의 추이를 보여주고 보고서를 생성해주지만 본 프로젝트(blabla)에서는 이 기능들은 구현하지 않는다. 그리고 다른 서비스들은 특별히 분석 대상을 제한하지 않지만, 본 프로젝트는 분석 대상을 영화로 제한한다. 하지만 본 프로젝트에서는 검색 대상만이 아니라 대상의 속성에 대한 선호도도 분석할 예정이다. 이런 기능을 포함하는 서비스는 오직 BEOLTONG 뿐이다.

### 3. 프로젝트 설계 내용

#### 3-1. 주요 기능

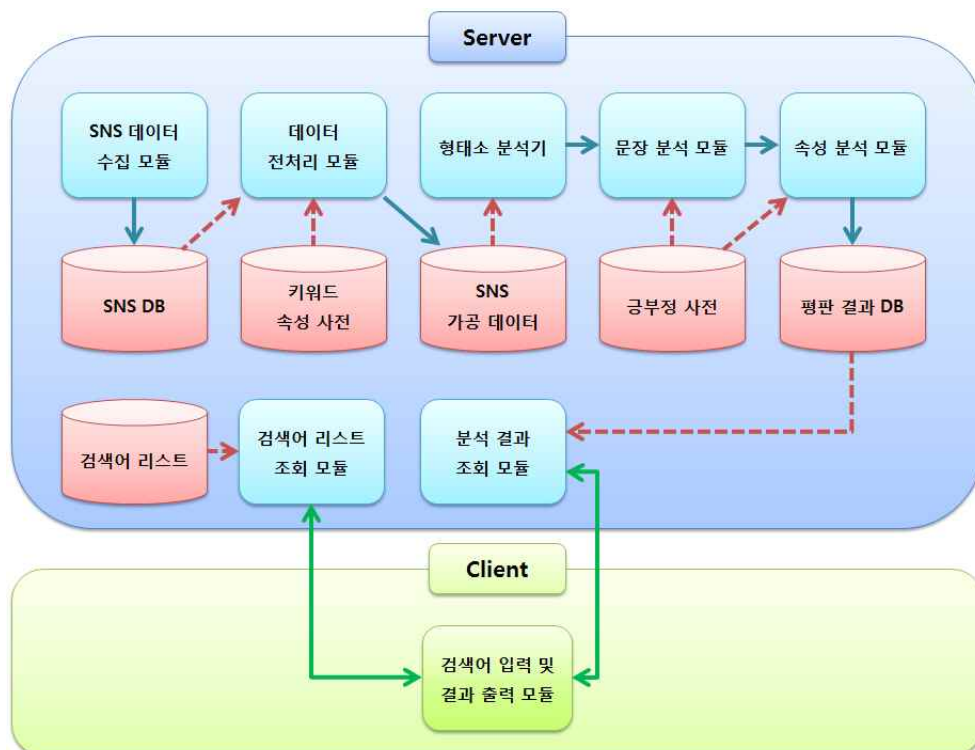
<1단계>

- 사용자가 입력한 영화에 대한 SNS상의 의견들을 출력하는 기능
- 사용자가 입력한 영화에 대한 SNS상의 대중의 선호도를 분석해 도표·그래프를 통해 시각화 할 수 있는 기능

<2단계>

- 사용자가 입력한 영화가 가진 속성에 대한 선호도 분석 기능 (예. 타이타닉 - 스토리, 음악, 연기력 등)
- 영화별 형용사 순위보기 기능
- 형용사별 영화순위보기 기능
- 전자제품, 주식에 관한 분석

#### 3-2. 모듈별 기능 명세서 및 시스템 흐름도



[그림 1] SNS 데이터 수집 모듈 세부 구성도



본 프로젝트에서 제시하는 시스템의 구조는 [그림 1]과 같다. 서버는 SNS 데이터를 수집하고 평판 분석을 한 후 그 결과를 제공하는 역할을 하며, 클라이언트는 서버에 평판 분석 결과를 요청해 웹페이지 및 안드로이드 애플리케이션으로 그 내용을 보여준다.

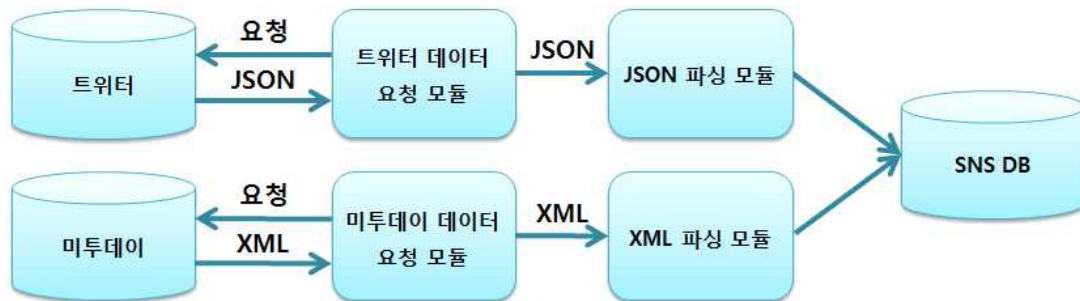
### 3-2-1. 클라이언트(Client)

- 검색어 입력 및 결과 출력 모듈 : 검색어를 입력 받고 분석된 데이터를 사용자에게 출력하는 모듈
- 클라이언트 클래스

Class	WebPage			각 영화의 분석 결과를 받아와
	보여주는 Class			
함수이름	인자타입	반환타입	설명	
addedToStarPageHandler	Event	void	프로그램이 시작될 때 실행된다. requestKeyObjectList()를 호출하는 함수	
keyObjectListResultHandler	ResultEvent	void	requestKeyObjectList()에 대한 응답으로 서버가 보낸 영화의 리스트와 각 영화의 이미지 URL과 영화 순위를 받고 그 내용을 화면에 띄우는 함수	
DisplayDefaultImage	void	void	메인 페이지의 영화 포스터 리스트를 화면에 나타내는 함수	
analyzedOpinionResultHandler	ResultEvent	void	requestAnalyzedOpinion()에 대한 응답으로 서버가 보낸 영화의 분석 결과를 화면에 나타내는 함수	
ErrorHandler	FaultEvent	void	HTTP 연결에 에러가 발생하면 에러 메시지를 띄우는 함수	
btnBackClickListener	MouseEvent	void	뒤로 버튼을 누르면 메인 페이지로 돌아가는 함수	
next_click	MouseEvent	void	메인 페이지에서 화살표를 누르면 영화 리스트의 내용이 바뀌게 하는 함수	
movie_clickHandler	MouseEvent	void	메인 페이지에 있는 영화 리스트의 포스터 하나를 클릭하면 그 영화의 분석 결과를 보여주는 페이지로 넘어가게 하는 함수	
localizeDateTime	String	String	날짜와 시각을 나타내는 문자열을 한국으로 만드는 함수	
movieNoteItemClickListener	int	void	분석 결과 페이지에 있는 영화 리스트의 영화 제목 하나를 클릭하면 그 영화의 분석 결과를 보여주는 페이지로 넘어가게 하는 함수	
changeStar	String	void	메인 페이지에서 분석 결과 페이지로 넘어가는 등의 페이지 변환을 수행하는 함수	


### 3-2-2. 서버(Server)

- SNS 데이터 수집 모듈



[그림 2] SNS 데이터 수집 모듈 세부 구성도

[그림 2]는 SNS 데이터 수집 모듈의 세부 구성도이다. 트위터와 미투데이에서 제공하는 API를 사용하여 입력한 키워드에 대한 데이터를 수집한다. 트위터는 JSON 형식, 미투데이는 XML 형식으로 데이터를 제공하기 때문에 각각의 형식에 맞는 파싱을 수행하고, 파싱 결과를 SNS DB에 삽입하는 모듈로 구성되어 있다. 아래 [그림 3]은 API를 사용한 트위터로의 요청에 대한 JSON 파일의 세부 내용이다.

```
{
  "completed_in": 0.041,
  "max_id": 216744931107151872,
  "max_id_str": "216744931107151872",
  "next_page": "?page=2&max_id=216744931107151872&q=batman&rpp=100",
  "page": 1,
  "query": "batman",
  "refresh_url": "?since_id=216744931107151872&q=batman",
  "results": [
    {
      "created_at": "Sun, 24 Jun 2012 04:09:53 +0000",
      "from_user": "flwmirelac",
      "from_user_id": 556880476,
      "from_user_id_str": "556880476",
      "from_user_name": "Mirela Carvalho",
      "geo": null,
      "id": 216744931107151872,
      "id_str": "216744931107151872",
      "iso_language_code": "pt",
      "metadata": {
        "result_type": "recent"
      },
      "profile_image_url": "http://a0.twimg.com/profile_images/2335387650/jw64zvlcrcc12fq58fwv_normal.jpeg",
      "profile_image_url_https": "https://si0.twimg.com/profile_images/2335387650/jw64zvlcrcc12fq58fwv_normal.jpeg",
      "source": "<a href='\"http://twitter.com/\">web</a>",
      "text": "RT @MG10_: e a mirela Wu00e9 o batman kkkk",
      "to_user": null,
      "to_user_id": 0,
      "to_user_id_str": "0",
      "to_user_name": null
    },
    {
      "created_at": "Sun, 24 Jun 2012 04:09:53 +0000",
      "from_user": "reporterocoz",
      "from_user_id": 184574559,
      "from_user_id_str": "184574559",
      "from_user_name": "Tuitero Inc Wu00f3modo",
      "geo": null,
      "id": 216744930813558784,
      "id_str": "216744930813558784",
      "iso_language_code": "es",
      "metadata": {
        "result_type": "recent"
      },
      "profile_image_url": "http://a0.twimg.com/profile_images/2236654470/Repor_C3_B1ero_20Twitter_20y_20Facebook_20_3_normal.JPG",
      "profile_image_url_https": "https://si0.twimg.com/profile_images/2236654470/Repor_C3_B1ero_20Twitter_20y_20Facebook_20_3_normal.JPG",
      "source": "<a href='\"http://blackberry.com/twitter\" rel='\"nofollow\">Twitter for BlackBerryWu00ae</a>",
      "text": "No es quien eres. Tus actos son los que te definen. #frasesdepeliculas (Batman)",
      "to_user": null,
      "to_user_id": 0,
      "to_user_id_str": "0",
      "to_user_name": null
    }
  ]
}
```

[그림 3] 수신한 트위터 JSON 세부 내용

트위터에서 제공하는 API를 사용하여 키워드로 "batman"을 사용한 결과, 수신한 JSON 파일이다. 상단의 첫 번째 구역은 파일의 헤더로서 요청에 따른 수행 시간, 페이지, 키워드, 쿼리 등으로 이루어

어져 있다. 두 번째 구역부터는 실제 트위터 사용자가 키워드 "batman"에 관하여 게시한 트윗 정보로 구성된다. 각각의 트윗 정보에는 게시 날짜, 사용자 ID, 사용자 이름, 텍스트 등으로 이루어진다.

#### - 수집 클래스

Class	TwitterCrawling
설명	Twitter에서 데이터를 수집하는 Class

함수이름	인자타입	반환타입	설명
executeSearch_click Handler	Event	void	TwitterAPI를 사용하여 JSON데이터를 불러오는 함수
completehandler	Event	void	JSON데이터를 DB테이블에 삽입하는 함수
handleConnected	Event	void	DB에 연결이 되었을 때 프로그램을 사용가능하게 만드는 함수

Class	ShowTest
설명	Me2day에서 데이터를 수집하는 Class

함수이름	인자타입	반환타입	설명
colorpicker1_clickHandler	Event	void	Me2dayAPI를 사용하여 XML데이터를 불러오는 함수
completehandler	Event	void	XML데이터를 DB테이블에 삽입하는 함수
handleConnected	Event	void	DB에 연결이 되었을 때 프로그램을 사용가능하게 만드는 함수

#### - 데이터 전처리 모듈



[그림 4] 데이터 전처리 모듈 세부 구성도

데이터 전처리 모듈은 [그림 4]와 같은 세부 모듈들로 이루어져있다. 우선, 평판 분석 결과의 신뢰성을 향상시키기 위해 광고 제거 모듈에서 광고 문장을 판단하여 제거한다. 광고 제거 과정에서는 여러 번 올라온 동일한 내용의 문서들을 제거한다. SNS에 여러 번 올라온 동일한 내용의 문서들을 살펴보면 높은 확률로 광고 문서라는 것을 알 수 있다. 이러한 문서들을 제거하기 위해서 새로운 문서가 추가될 때마다 해시(hash) 값이 같은 문서를 찾아 내용이 완전히

같은 경우 광고 문서로 판단해 광고 문서 테이블에 저장하고 입력 데이터에서 삭제한다. 그리고 띄어쓰기 교정 모듈에서는 형태소 분석이 원활히 수행될 수 있도록 문장의 띄어쓰기를 교정하며, 마지막으로 문장에서 속성들을 추출하기 위한 속성 추출 모듈이 있다.

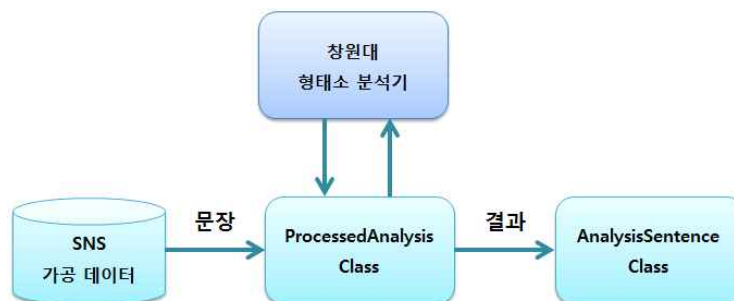
속성 추출 모듈은, SNS데이터와 DB에 저장되어 있는 속성과의 비교를 통해 속성을 추출한다. 그리고 광고 제거 모듈은 글의 Hash값을 DB에 저장하고 Hash값이 같은 것이 발견되면 flag에 체크하고 flag가 체크 되지 않은 Hash값과 같은 데이터는 삭제 후, advertise 테이블에 임시 저장한다. 띄어쓰기 교정 모듈은 직접 개발한 것을 사용하는 것이 아니라 국민대학교의 강승식 교수가 개발한 형태소 분석기에 포함되어 있는 띄어쓰기 교정 함수를 사용한다. [그림 5]는 띄어쓰기 교정 모듈의 간단한 사용 예제이다.



[그림 5] 띄어쓰기 모듈 사용 예제

#### - 형태소 분석기

형태소 분석기는 설계 및 개발 자체만으로도 상당한 시간을 필요로 하는 하나의 프로젝트가 될 수 있기 때문에 본 프로젝트에서는 창원대에서 제공하는 형태소 분석기를 사용한다. 창원대 형태소 분석기는 자바에서 사용 가능하여 문자열을 다루기 쉽고 상세한 매뉴얼을 제공하기 때문에 형태소 분석기에서 제공하는 다양한 API를 손쉽게 사용한 장점이 있다. 창원대 형태소 분석기 사용 예는 [그림 6]과 같다.



[그림 6] 형태소 분석기 사용 예제

창원대 형태소 분석기는 기본적으로 감성 문장 분석을 위해 문장을 체언, 용언, 수식언으로 구분하기 위해 사용된다. 먼저 SNS 가공 데이터 DB로부터 문장을 가져온 뒤, 창원대 형태소 분석기에서 제공하

는 API를 사용하여 문장의 형태소 분석을 수행한다. AnalysisSentence 클래스는 문장의 분석 결과로서 식별번호와 선호도 수치 그리고 ArrayList로 선언된 명사, 동사, 형용사, 관형사, 부사를 각각 저장한다.

★ API 설명

```
String tagger::POSTagger(
String input,          /* 입력 문장 */
TaggerConf conf);     /* 태거 환경 설정 클래스 TaggerConf */
```

입력문장과 태거 환경 설정 클래스를 인자로 받아 결과 값을 String으로 출력한다.

[예제 소스]

```
String input = new String("결과 보고서를 위한 예제결과입니다.");
TaggerConf taggerConf =
new TaggerConf("./dic/LEXICON.morph2.idx", "./dic/LEXICON.morph2.dat",
               "./dic/LEXICON.sym2.idx", "./dic/LEXICON.sym2.dat", "./dic/LEXICON.prob.mhdict");
ret = tagger.POSTagger(input, taggerConf);
System.out.println(ret);
```

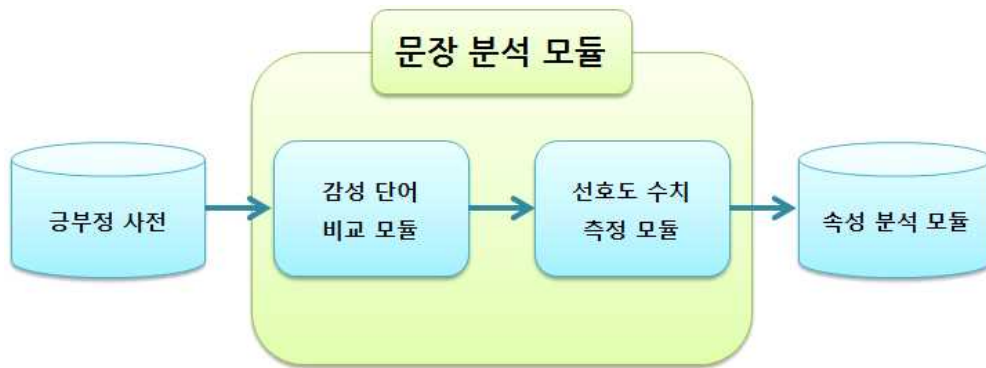
[예제 결과]

```
결과보고서를      BOS      결과보고서/NNG+를/JKO
위한              위하/VV+ㄴ/ETM
예제결과입니다.   EOS      예제결과/NNG+이/VCP+ㅂ니다/EF+./SF
```

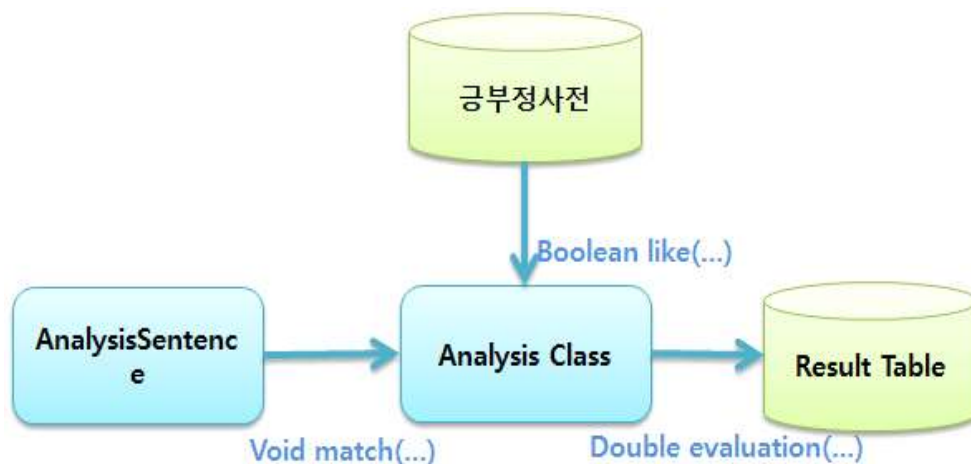
BOS는 분석 문장의 시작을 의미하고 EOS는 분석 문장의 끝을 의미한다. 그 외 '/' 이후에 나오는 알파벳 문자의 의미는 아래 [표 2]을 참조한다.

대분류	소분류	세분류	대분류	소분류	세분류
(1) 체언	명사 NN	일반명사 NNG	(3) 수식언	관형사 MM	관형사 MM
		고유명사 NNP		부사 MA	일반부사 MAG
		의존명사 NNB	(4) 독립언	감탄사 IC	접속부사 MAJ
	대명사 NP	대명사 NP	(5) 관계언	격조사 JK	주격조사 JKS
	수사 NR	수사 NR			보격조사 JKC
(2) 용언	동사 VV	동사 VV			관형격조사 JKG
	형용사 VA	형용사 VA			목적격조사 JKO
	보조용언 VX	보조용언 VX			부사격조사 JKB
	지정사 VC	긍정지정사 VCP			호격조사 JKV
		부정지정사 VCN			인용격조사 JKQ
				보조사 JX	보조사 JX

[표 2] 품사 태그표



[그림 7] 문장 분석 모듈 세부 구성도



[그림 8] 문장 분석 과정

[그림 7]의 문장 분석 모듈은 본 프로젝트에서 가장 핵심적인 기능을 수행하는 모듈로서, 형태소 분석기를 통해 얻은 문장의 성분(주어, 동사, 부사 등)들로부터 문장의 선호도 수치를 계산하는 모듈이다. 문장 분석 모듈은 세부 모듈인 감성 단어 비교 모듈과 선호도 수치 측정 모듈로 이루어져 있다. 문장 분석 과정은 [그림 8]과 같다.

먼저 감성 단어 비교 모듈에서는 해당 문장에 미리 구축한 금부정 사전의 단어가 포함되어 있는지 확인한다. 금부정 사전에 단어를 어미가 제거된 형태로 저장해둔 다음, 금부정 사전의 모든 단어와 형태소 분석 과정을 거친 문장의 각 형태소의 앞부분이 일치하는지 여부를 확인한다. 예를 들어 문장에 ‘떨떠름하네염’이라는 단어가 포함되어있는 경우 형태소 분석기가 완벽히 분석하지 못하고 ‘떨떠름하네염’이 하나의 형태소라고 분석할 수 있는데 이런 경우 사전의 ‘떨떠름’이란 단어와 앞부분이 일치하므로 ‘떨떠름하네염’은 ‘떨떠름’의 의미라고 생각할 수 있다. 만약 금부정 사전에  $n$ 개의 단어가 있고 문장에 포함된 형태소가  $m$ 개면 문장 내의 모든 감성 단어를 찾는 데는  $\theta(n \times m)$ 의 시간이 소요된다.

선호도 수치 측정 모듈에서는 각 문장을 긍정적인 문서와 부정적인 문서로 분류한다. 일치하는 감성 단어가 있을 경우 선호도 수치 측정 모듈에서는 발견한 감성 단어 각각의 미리 정의된 금부정

값을 모두 더해 문장이 긍정적인지 부정적인지를 판별하고 그 강도를 계산한다. 본 프로젝트에서는 극성의 분류와 강도를 구하기 위해 [수식 1]을 제안한다. 집합  $W(d)$ 는 문서  $d$ 에서 추출된 형태소의 집합  $A(d)$ 와 감성 단어 집합  $E$ 의 교집합이다. 그리고  $word_{d,k}$ 는 집합  $W(d)$ 의 원소이며,  $|W(d)|$ 는 문서  $d$ 에 포함된 감성 단어의 수,  $posneg()$ 는 단어의 쾌-불쾌 수치를 나타내는 함수,  $active()$ 는 단어의 활성화 수치를 나타내는 함수,  $pref(d)$ 는 문서  $d$ 의 선호도를 나타내는 함수이다.  $a$ 는 임계값으로 쾌-불쾌 수치가  $a$ 보다 크면 긍정을, 작으면 부정을 의미한다.

$$pref(d) = \sum_{k=1}^{|W(d)|} ((posneg(word_{d,k}) - \alpha) \times active(word_{d,k}))$$

$$word_{d,k} \in W(d) = A(d) \cap E$$

[수식 1] 문서의 극성 분류 및 극성의 강도 계산식  
( $\alpha = 3.81$ )

#### - 문장분석 클래스

Class	isWhat
설명	태그가 문장성분의 어디에 속하는지 알려주는 함수를 가진 Class

함수이름	인자타입	반환타입	설명
MA	String	Boolean	입력된 String태그가 부사면 True를 아니면 False를 return하는 함수
MM	String	Boolean	입력된 String태그가 관형사면 True를 아니면 False를 return하는 함수
N	String	Boolean	입력된 String태그가 명사면 True를 아니면 False를 return하는 함수
VA	String	Boolean	입력된 String태그가 형용사면 True를 아니면 False를 return하는 함수
VV	String	Boolean	입력된 String태그가 동사면 True를 아니면 False를 return하는 함수

Class	ScoreString
설명	하나의 문장 성분과 문장 성분의 감정수치를 가지고 있는 Class

속성명	속성타입	속성설명
component	String	문장 속의 문장성분(명사, 형용사, 동사, 부사, 관형사 등)
score	double	문장성분의 감정수치

Class	AnalysisSentence
설명	문장의 분석된 결과를 가지고 있는 Class



속성명	속성타입	속성설명
SentenID	int	문장의 식별번호
Noun	ArrayList<ScoreString>	문장을 구성하는 명사의 집합
VV	ArrayList<ScoreString>	문장을 구성하는 동사의 집합
VA	ArrayList<ScoreString>	문장을 구성하는 형용사의 집합
MM	ArrayList<ScoreString>	문장을 구성하는 관형사의 집합
MA	ArrayList<ScoreString>	문장을 구성하는 부사의 집합
total	double	문장 성분들이 가진 문장 수치의 총합

Class	ProcessedAnalysis
설명	형태소분석기의 API를 이용하여 문장을 문장성분별로 구분하는 Class

함수이름	인자타입	반환타입	설명
getTagger	String	AnalysisSentence	문장을 입력값으로 받아 창원대형태소분석기를 통해 문장의 성분들을 AnalysisSentence의 형태로 반환하여 return하는 함수

Class	Getsearch
설명	DB에 접속하여 현재 상영 중인 영화명을 가져오는 Class

속성명	속성타입	속성설명
item	String[]	결과값인 영화명이 저장되어지는 String배열

함수이름	인자타입	반환타입	설명
Getsearch			Getsearch Class의 생성자로 DB에 연결을 한 후 영화명을 item변수에 저장하는 함수
getItem	void	String[]	item String의배열값을 반환하는 함수

Class	Analysis
설명	AnalysisSentence의 각 문장성분을 긍부정사전과 비교하여 감성문장인지 찾아낸다. 각 문장성분의 감정수치를 찾은 후 문장의 총 감정수치를 계산하는 Class

함수이름	인자타입	반환타입	설명
like	String[], String	Boolean	문장성분이 긍부정사전에 존재하는지 검사하는 함수
match	AnalysisSentence, String[], double[]	void	문장성분별로 like함수를 호출해 긍부정사전에 존재하는지 검사하고 존재하면 해당 감정수치를 AnalysisSentence변수에 할당하는 함수
evaluation	AnalysisSentence	double	문장에서 문장성분들의 감정수치의합을 의미하는 total값을 return하는 함수
getMe2Id	void	String[]	Me2day 테이블에서 해당 Me2day문장의 ID를 가져오는 함수
getMe2Sentence	void	String[]	Me2day의 테이블에서 문장을 가져오는 함수
getTwitID	void	String[]	Twitter의 테이블에서 해당 Twitter문장의 ID를 가져오는 함수
getTwitSentence	void	String[]	Twitter의 테이블에서 문장을 가져오는 함수
inputResultMe2day	double, int, String	void	Me2day의 분석된 결과를 DB테이블에 집어넣는 함수
inputResultTwitter	double, int, String	void	Twitter의 분석된 결과를 DB테이블에 집어넣는 함수
adjective_match	String, int	void	감성 문장 내에서 형용사를 추출하는 함수
insert_adjective	void	void	같은 영화에 같은 형용사가 2회 이상 나타난 경우 DB에 영화명과 함께 저장하는 함수
		void	감성단어를 DB에서 읽어 들여 배열에 저장하는 함수
			속성을 불러와 배열에 저장하고 Mapping 되어
setAttr	void	void	있는 속성을 불러와 추가적으로 배열에 저장하는 함수
getMe2Sentence	void	void	me2day 데이터를 불러와 like, match, evaluation을 수행하는 함수
setAttrMe	void	void	me2day 데이터를 불러와 속성에 대하여 like, match, evaluation을 수행하는 함수
getTwitSentence	void	void	twitter 데이터를 불러와 like, match, evaluation을 수행하는 함수
setAttrTwitSentence	void	void	twitter 데이터를 불러와 속성에 대하여 like, match, evaluation을 수행하는 함수
inputAttr	double, int, String	void	twitter데이터를 속성에 관해 분석한 후, DB에 그 결과를 저장하는 함수

함수이름	인자타입	반환타입	설명		
inputAttrResultTwitt er	double, int, String, String	void	me2day 데이터를 속성에 관해 분석 한 후, DB에 그 결과를 저장하는 함수		
inputNum Result	void	void	Result 테이블에 저장되어 있는 값을 계산해서 하나의 테이블에 저장하는 함수		
inputAttrN umResult	void	void	Result 테이블에 저장되어 있는 속성 이 저장되어 있는 값을 계산해서 하 나의 테이블에 저장하는 함수		
attrLike	String, String, String	boolean	문장에 속성이 있는지 검색하는 함수		
clearResul t	void	void	result 테이블을 삭제하는 함수		
clearNum Result	void	void	num_result 테이블을 삭제하는 함수		
clearAdjec tive	void	void	adjective 테이블을 삭제하는 함수		

- 속성 분석 모듈 : 데이터 전처리를 과정을 통해 추출된 속성들을 참조하여 해당 속성이 문장 내에서  
긍정적인지 부정적인지를 판단하는 모듈
- 긍부정 사전

감성 단어와 해당 감성 단어의 긍부정 값이 저장되어 있는 사전이다. 긍부정 사전 구축에는 [9]의 부록 1에 수록된 단어들의 쾌-불쾌 수치와 활성화 수치를 사용한다. 긍부정 사전 구축에 사용된 단어의 예는 [표 3]과 같다. 쾌-불쾌 수치는 단어가 가진 긍정 또는 부정적인 느낌의 정도를 나타내고 활성화 수치는 단어가 가진 느낌이 얼마나 강렬한지를 나타낸다. 쾌-불쾌 수치는 부정적인 느낌이 강한 단어일수록 0에 가깝고 긍정적인 느낌이 강한 단어일수록 7에 가깝다.

단어	쾌-불쾌	활성화
노엽	2.48	5.32
매혹	4.84	4.59
뿌듯	5.75	4.63
스산	2.85	3.72
애석	2.67	3.78
우수	4.28	3.38
쾌감	5.79	5.91
처참	1.64	4.5

[표 3] 감성 단어 구축에 사용된 단어 예

그리고 SNS에서 사용되는 단어들 중 [9]의 부록 1에 없는 단어들을 조사해 정리한 다음, 설문조사를 통해 수치를 결정해 긍부정 사전에 추가하여 사용하였다. 설문조사는 [9]의 부록 1의 단어들 중 샘플을 뽑아 추가할 단어들이 샘플로 뽑힌 단어들 중 어느 단어와 비슷한 느낌을 가졌는지 선택하게 하는 방식으로 수행되었다. 추가되는 단어들은 샘플로 뽑힌 단어들에 매핑시켜 해당 수치를 사용하였다. 매핑된 단어의 예는 [표 4]와 같다. 매핑 테이블과 긍부정 사전의 매핑 예시는 [그림 9]와 같다. 매핑 테이블의 WORD 속성은 새로 추가된 감성 단어를 의미하고, DIC\_ID 속성은 긍부

정 사전 내의 매핑될 WORD의 ID 값을 가진다.

단어	감성 단어 사전의 단어
빽	가엎
안습	가엎
상당	부럽
유쾌	공감
감사	공감
엄청	행복
인상	공감
괜차	부럽

[표 4] 매핑된 단어의 예

ID	WORD	래-불래	활성화	ID	DIC_ID	WORD
1	감동	5.45	4.47	1	1	두근두근
2	불래	2.13	5.46	2	1	강추
3	슬프	2.69	3.44	3	2	안습

<감성 단어 사전>                      <매핑 테이블>

[그림 9] 감성 단어 사전 및 매핑 테이블

- 평판 결과 DB : 평판 분석이 수행된 문장의 선호도 수치를 저장하는 사전
- 분석 결과 조회 모듈 : 검색어를 인자로 받아 평판 결과 DB에서 분석 결과를 조회하여 검색어 입력 및 결과 출력 모듈에게 분석 결과를 전송하는 모듈

Class	KeyObjectList
설명	현재 분석 가능한 영화의 리스트와 각 영화의 이미지 URL을 돌려주는 Class

함수이름	인자타입	반환타입	설명
doGet	HttpServletRequest, HttpServletResponse	void	현재 분석 가능한 영화의 리스트와 각 영화의 이미지 URL을 돌려주는 함수

Class	AnalyzedOpinion
설명	요청한 영화의 분석 결과를 돌려주는 Class

함수이름	인자타입	반환타입	설명
doGet	HttpServletRequest, HttpServletResponse	void	요청한 영화의 분석 결과를 돌려주는 Class

Class	AdjectiveMovieRanking
설명	요청한 형용사에 대한 영화 순위를 돌려주는 Class

함수이름	인자타입	반환타입	설명
doGet	HttpServletRequest, HttpServletResponse	void	요청에 대한 형용사별 영화 순위를 돌려주는 Class

Class	MySQLConnection
설명	DBMS인 MySQL과의 연결을 관리하는 Class

함수이름	인자타입	반환타입	설명
isMySQLDriverExist	void	Boolean	MySQL 드라이버가 존재하는지 확인하는 함수
getConnection	void	Connection	요청한 모듈에게 이미 생성되어 있는 MySQL과의 연결을 반환하는 함수

### 3-3-1. E-R 다이어그램



20

### 3-3-2. 테이블 명세서

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
movie_name	id	int	not null	○			
	name	varchar	null				
	thumbnail	varchar	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
movie_attribute	id	int	not null	○			
	attribute	varchar	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
result	id	int	not null	○			
	posnav	double	null				
	twitter_id	int	null		○	twitter	id
	emotional_flag	int	not null				
	me2day_id	varchar	null		○	me2day	id
	movie_name	varchar	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
num_result	id	int	not null	○			
	pos_sum	double	null				
	nav_sum	double	null				
	movie_name	varchar	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
me2day	id	int	not null	○			
	value	varchar	null				
	movie_name	varchar	null				
	date	varchar	null				
	flag	int	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
twitter	id	int	not null	○			
	value	varchar	null				
	movie_name	varchar	null				
	date	varchar	null				
	flag	int	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
processed _me2day	me2day_id	int	null		○	me2day	id
	id	int	not null	○			
	me2day	varchar	null				
	movie_name	varchar	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
processed _twitter	twitter_id	int	null		○	twitter	id
	id	int	not null	○			
	me2day	varchar	null				
	movie_name	varchar	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
general_di ctionary	id	int	not null	○			
	text	varchar	null				
	real	double	null				
	friendly	double	null				
	posnav	double	null				
	active	double	null				



테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
mapping_t able	id	int	not null	○			
	dic_id	int	null		○	general_di ctionary	id
	text	varchar	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
adverb_di ctionary	id	int	not null	○			
	text	varchar	null				
	value	double	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
ending_di ctionary	id	int	not null	○			
	text	varchar	null				
	value	double	null				



테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
attribute_ mapping_t able	id	int	not null	○			
	attr_id	int	null		○	attribute	id
	value	varchar	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
actor_map ping_table	id	int	not null	○			
	movie_name	varchar	null				
	actor_name	varchar	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
num_adjective	id	int	not null	○			
	movie_name	varchar	null				
	adjective	varchar	null				
	num_adjective	int	null				

테이블이름	열 이름	데이터형식	Null 유무	기본키	외래키	FK 테이블이름	FK 열 이름
movie_ranking	id	int	not null	○			
	movie_name	varchar	null				
	ranking	int	null				
	attribute	varchar	null				

### 3-3-3. 테이블 설명

Adverb_dictionary				Movie_attribute		
blabla.adverb_dictionary: 4 rows total				blabla.movie_attribute: 6 rows total		
	id	text	value		id	attribute
	1	하지만	0.5		1	연출
	2	매우	1.5		2	구성
	3	정말	1.5		3	줄거리
	4	반면에	0.5		4	출연진
					5	감독
					6	음향

Movie_name		
movie_name: 7 rows total		
id	name	thumbnail
3	차형사	http://movie.phinf
2	맨인블랙3	http://movie.phinf
4	내아내의모든것	http://movie.phinf
5	미확인동영상	http://movie.phinf
6	머벤저스	http://movie.phinf
7	돈의맛	http://movie.phinf
8	후궁	http://movie.phinf

[그림 11] Adverb\_dictionary, Movie\_attribute, Movie\_name 테이블

[그림 11]에서의 Adverb\_dictionary는 부사어를 저장하는 어휘 사전이다. 중립값 1을 기준으로 순접의 부사어인 경우 문장의 긍부정 강도를 높이도록 1이상의 값을 곱해주고, 역접의 부사어인 경우는 문장의 긍부정 강도를 낮추도록 1이하의 값을 곱해준다. movie\_attribute 테이블은 영화의 속성에 대하여 저장한 테이블이다. movie\_name 테이블은 영화명과 각 영화의 thumbnail의 링크주소를 저장하고 있다. thumbnail을 이용해 포스터를 불러온다.

## me2day

me2day: 12,434 rows total, limited to 1,000					Next	Show all	Sorting	Columns	Filter
id	value	date	movie_name	flag					
13253	후궁 제왕의 첩을 보고...	2012-06-05T01:32:55+0900	후궁	1					
13252	[후궁성형술, 활기찬병원] 활기찬병원에서 알려드리...	2012-06-04T17:26:18+0900	후궁	1					
13251	후궁	2012-06-05T18:26:35+0900	후궁	1					
13250	투머코리아 - [6월 개봉영화] '후궁: 제왕의 첩 6월 판...	2012-06-04T18:46:47+0900	후궁	1					
13249	후궁:제왕의 첩	2012-06-05T14:26:07+0900	후궁	1					
13248	[까칠미소식] '예로탁 사극'이라더니 미련 걸 숨겨 ...	2012-06-04T19:45:47+0900	후궁	1					
13247	후궁 무대인사 정보	2012-06-05T12:58:57+0900	후궁	1					
13246	&#39;후궁&#39;; 조여정 '마지막 정사신만 4일, 녹초...	2012-06-04T20:30:56+0900	후궁	1					
13245	[후궁:제왕의 첩]대전도시철도공사와 함께하는 시사...	2012-06-05T02:19:55+0900	후궁	1					
13244	오 후궁 개봉했네! 25:30 예약- n n	2012-06-05T11:02:16+0900	후궁	1					
13243	보고싶은 영화 후궁 롯데백화점 탈모	2012-06-05T13:41:09+0900	후궁	1					
13242	'후궁' 임권택 감독과 최고 스타들의 만장일치 찬사'...	2012-06-05T10:15:22+0900	후궁	1					
13241	잔인한 공중권력다툼 속 살기 위해 처절히 변해마했...	2012-06-09T14:20:12+0900	후궁	1					

## twitter

twitter: 10,500 rows total, limited to 1,000					Next	Show all	Sorting	Columns	Filter
id	value	date	movie_name						
10512	후궁?	Mon, 11 Jun 2012 08:41:24 +0000	후궁						
10511	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화 트...	Mon, 11 Jun 2012 08:41:32 +0000	후궁						
10510	영화 후궁보고 예머컨도 약한 코엑스 방황하며 고마...	Mon, 11 Jun 2012 08:45:10 +0000	후궁						
10509	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화를 ...	Mon, 11 Jun 2012 08:46:00 +0000	후궁						
10508	하루 종일 영화홀릭. 후궁+건축학개론. 그리고 내일 ...	Mon, 11 Jun 2012 08:48:03 +0000	후궁						
10507	RT @megabox_event: [무비출나잇 금주상영작] 6/8(...	Mon, 11 Jun 2012 08:48:13 +0000	후궁						
10506	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화를 ...	Mon, 11 Jun 2012 08:49:56 +0000	후궁						
10505	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화를 ...	Mon, 11 Jun 2012 08:51:05 +0000	후궁						
10504	RT @movie0free: 프롬메테우스랑 후궁을 보고난 후에...	Mon, 11 Jun 2012 08:51:40 +0000	후궁						
10503	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화를 ...	Mon, 11 Jun 2012 08:53:31 +0000	후궁						
10502	후궁:제왕의 첩 - 광기어린 사랑과 달콤한 욕망의 끝...	Mon, 11 Jun 2012 08:55:32 +0000	후궁						
10501	후궁 VIII정선카지노비 http://t.co/vCkhLdaG[[[⊙.⊙=&...	Mon, 11 Jun 2012 08:56:09 +0000	후궁						

[그림 12] me2day, twitter 테이블

[그림 12]에서의 me2day 테이블은 미투데이의 미투와 글이 작성된 날짜, 작성자id와 thumbnail을 저장한다. 그리고 twitter 테이블은 트위터의 멘션과 멘션이 작성된 날짜, 작성자id와 thumbnail을 저장한다. 현재 me2day 테이블에는 12434개의 레코드가 있고, twitter테이블에는 10500개의 레코드가 존재하는 상태이다. 그리고 me2day, twitter테이블에 존재하는 flag 필드는 띄어쓰기를 했는지 안했는지에 관한 상태를 저장하는 필드이다. 띄어쓰기가 된 문장은 flag가 1이 된다. 그리고 date 필드의 비교를 통해 중복된 데이터를 처리한다.

## Processed\_me2day

processed\_me2day: 10,438 rows total, limited to

me2day_id	id	me2day	movie_name
13253	10438	후궁 제왕의 첩을 보고...	후궁
13252	10437	[후궁성형술, 활기찬 병원] 활기찬 병원에서 알려드...	후궁
13251	10436	후궁	후궁
13249	10435	후궁:제왕의 첩	후궁
13248	10434	[까칠이 소식] `에로틱 사극`이라더니 이런 걸 숨겨 ...	후궁
13247	10433	후궁 무대인 사 정보	후궁
13246	10432	&#39;후궁 &#39; 조여 정 "마지막 정사신만 4일, 녹...	후궁
13245	10431	[후궁:제왕의 첩]대전도시 철도 공사와 함께하는 시...	후궁
13244	10430	오 후궁 개봉했네! 25:30 예약 - n n	후궁
13243	10429	보고싶은 영화 후궁 롯데백화점 탈모	후궁
13242	10428	`후궁`임권택 감독과 최고 스타들의 만장 일치 찬사`...	후궁
13241	10427	잔인한 궁중권력 다툼 속 살기 위해 처절히 변해마...	후궁

## Processed\_twitter

processed\_twitter: 10,080 rows total, limited to 1,000

twitter_id	id	twitter	movie_name
10512	10080	후궁?	후궁
10511	10079	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화 트...	후궁
10510	10078	영화 후궁 보고 에 어쨌든 약한 코엑스 방황하며 고 ...	후궁
10509	10077	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화를 ...	후궁
10508	10076	하루 종일 영화홀릭. 후궁 +건축학개론. 그리고 내일...	후궁
10507	10075	RT @megabox_event: [무 비올나 잇 금주상영작] 6/8(...	후궁
10506	10074	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화를 ...	후궁
10505	10073	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화를 ...	후궁
10504	10072	RT @movie0free: 프로메테우스랑 후궁을 보고난 후에...	후궁
10503	10071	후궁 RT @AKMALLcom: RT와 함께 보고 싶은 영화를 ...	후궁
10502	10070	후궁:제왕의 첩 - 광기 어린 사랑과 달콤한 욕망의 끝...	후궁
10501	10069	후궁 V8정선카지노 비 http://t. co/vCkhLdaG[[[ @. @ =...	후궁

[그림 13] Processed\_me2day, Processed\_twitter 테이블

[그림 13]에서 processed\_me2day 테이블은 me2day 테이블에 존재하던 미투를 띄어쓰기 하여 새로 저장한 테이블이다. 따라서 me2day\_id를 foreign key로 가지고 me2day의 id와 연결된다. 마찬가지로 processed\_twitter 테이블은 twitter 테이블에 존재하던 멘션을 띄어쓰기 하여 새로 저장한 테이블이다. twitter\_id를 foreign 키로 가지고 twitter의 id와 연결된다.



## General\_dictionary

general\_dictionary: 434 rows total

id	text	real	friendly	posnav	active
434	회회낙락하다	4.16	4.17	4.71	4.8
433	회열	4.49	4.42	5.32	5.44
432	회비	3.76	4.08	3.8	4.61
431	회	3.65	3.22	5.21	4.61
430	홍취	3.4	3.07	4.71	5.08
429	홍분	4.92	5.61	4.44	6.02
428	홍미진진하다	5.07	5.61	5.4	5.21
427	홍마롭다	5.05	5.82	5.56	4.76
426	홍나다	4.73	4.54	5.5	6
425	홍겁다	5.52	5.09	5.72	5.5
424	홍죽하다	4.9	4.9	5.77	4.14
423	홍모하다	4.13	4.11	4.94	4.34
422	호뜻하다	5.26	5.46	5.48	3.85
421	후회하다	4.74	5.82	2.45	4.06
420	후련하다	5.44	5.3	5.33	3.61
419	회한	3.28	3.42	2.49	4.41
418	회오	2.35	1.7	3.63	4.1
417	회개	3.19	3.7	3.91	3.58
416	활솔하다	5.3	5.18	5.72	5.14
415	활송하다	3.63	3.58	4.07	4.18
414	활당하다	4.75	5.53	3.3	4.92
413	활공하다	3.51	3.67	4.25	4.57
412	환회	4.5	4.7	5.72	5.54
411	환멸	3.49	3.34	1.71	5.62
410	화나다	4.86	5.65	2.17	5.91
409	홀가분하다	5.71	5.18	6.24	3.26
408	혼비백산하다	3.44	3.79	2.32	5.98
407	호감	5.62	5.31	5.72	5.03
406	형오하다	4.35	5.04	1.8	5.57
405	허허롭다	2.88	2.69	3.07	3.06
404	허탈하다	4.36	4.99	2.67	3.2
403	충족감	4.26	4.58	5.27	4.15

## Mapping\_table

id	dic_id	text
119	63	간지나다
140	63	감사
142	63	감쇄
98	128	감지덕지
77	4	갑갑
154	63	강추
137	63	고상
136	63	고상하다
135	63	고상한
91	128	공감
106	128	편향다
173	391	구려
52	391	구리다
175	391	구린
174	391	구림
122	63	귀여운
121	63	귀엽다
127	63	기대
139	63	갈끔
85	4	깜깜
189	4	깜깜
57	391	꽤다
193	4	껌껌
172	391	꼴랐다
184	4	찍찍
61	4	질질거리다
108	128	낮다
65	4	단순
66	4	단조
123	63	달콤
96	128	당기다
103	128	대단하다

[그림 14] General\_dictionary, Mapping\_table 테이블

[그림 14]에서는 general\_dictionary 테이블과 mapping\_table이라는 테이블을 보여주는데, general\_dictionary 테이블은 긍부정을 나타내는 동사와 명사를 저장한다[9]. real 필드의 값은 원형성으로 감성문장의 올바른 예인지 아닌지를 0~7 사이의 값으로 구별하게 하는 수치이다. 원형성의 예를 들면 '몽롱하다, 든든하다' 등의 신체의 상태를 표현 하는 말은 감정을 표현하는 말이 아니라고 판단되기 때문이다. friendly 필드의 값은 사람들의 각 단어에 관한 친숙도를 0~7 사이의 값으로 나타낸다. friendly의 예를 들면, '회오, 송연하다'등은 친숙도가 낮은 단어가 된다. posnav 필드는 긍부정 정도를 0~7 사이의 값으로 나타낸다. active 필드는 0~7 사이의 값으로 활성도가 높고 낮음을 나타낸다. 여기서 활성도란 얼마나 감정이 평소 감정과 편차가 큰지 크지 않은지에 대해 그리고 사람이 느끼기에 얼마나 단어가 활발한 느낌을 주는지에 관한 내용부분이다. 예를 들면, '안락하다, 평온하다, 안도하다, 무력감'등이 가장 활성도가 낮았으며 '열광, 분노하다'와 같은 단어의 활성도가 높게 측정되었다. 현재는 friendly, real 값은 사용하지 않고 있으며 posnav와 active 값으로만 선호도를 계산한다. 그리고 mapping\_table은 SNS상에 있는 새로운 단어들을 추가한 테이블이다. 여기서 데이터를 추가한 방법은 설문조사를 통해 숭실대학교 컴퓨터학부 학생 54명의 쾌불쾌의 감정을 1에서 7사이의 값으로 설문지를 통해 선택을 받았으며 1에 대표되는 단어, 2에 대표되는 단어 등 각각의 수치에 해당하는 대푯값들과 연결한다. 따라서 dic\_id는 foreign\_key이고 general\_dictionary의 id와 연결된다.

id	posnav	twitter_id	emotional_flag	me2day_id	movie_name	attribute	flag	id	pos_sum	nav_sum	movie_name	attribute
1	-2.6973	(NULL)	1	1	다크나이트 라이즈	(NULL)	0	1	5240.976	-803.732000000001	다크나이트 라이즈	(NULL)
2	-4.7124	(NULL)	0	3	다크나이트 라이즈	(NULL)	0	11	63.7913	(NULL)	다크나이트 라이즈	연출
3	0	(NULL)	0	4	다크나이트 라이즈	(NULL)	0	21	(NULL)	-6.7823	다크나이트 라이즈	구성
4	0	(NULL)	0	5	다크나이트 라이즈	(NULL)	0	31	78.871	-19.2348	다크나이트 라이즈	스토리
5	0	(NULL)	0	6	다크나이트 라이즈	(NULL)	0	41	26.7542	-10.7369	다크나이트 라이즈	음악
6	0	(NULL)	0	7	다크나이트 라이즈	(NULL)	0	51	83.1429	-2.6973	다크나이트 라이즈	배우
7	-2.6973	(NULL)	1	8	다크나이트 라이즈	(NULL)	0	61	18.0352	(NULL)	다크나이트 라이즈	연기
8	0	(NULL)	0	9	다크나이트 라이즈	(NULL)	0	71	(NULL)	(NULL)	다크나이트 라이즈	화면
9	2.7528	(NULL)	1	10	다크나이트 라이즈	(NULL)	0	2	8281.002700000006	-1502.3924	도둑들	(NULL)
10	8.4852	(NULL)	1	11	다크나이트 라이즈	(NULL)	0	12	46.4253	-4.9909	도둑들	연출
11	0	(NULL)	0	13	다크나이트 라이즈	(NULL)	0	22	41.8354	-5.3946	도둑들	구성
12	0	(NULL)	0	14	다크나이트 라이즈	(NULL)	0	32	250.7365	-5.916	도둑들	스토리
13	0.8303	(NULL)	1	15	다크나이트 라이즈	(NULL)	0	42	15.2174	(NULL)	도둑들	음악
14	0	(NULL)	0	16	다크나이트 라이즈	(NULL)	0	52	400.907	-13.6598	도둑들	배우
15	0	(NULL)	0	17	다크나이트 라이즈	(NULL)	0	62	225.9033	(NULL)	도둑들	연기
16	0	(NULL)	0	18	다크나이트 라이즈	(NULL)	0	72	(NULL)	(NULL)	도둑들	화면
17	0	(NULL)	0	19	다크나이트 라이즈	(NULL)	0					
18	9.4248	(NULL)	1	20	다크나이트 라이즈	(NULL)	0					
19	0	(NULL)	0	21	다크나이트 라이즈	(NULL)	0					
20	0	(NULL)	0	23	다크나이트 라이즈	(NULL)	0					
21	5.3897	(NULL)	1	25	다크나이트 라이즈	(NULL)	0					
22	-7.4732	(NULL)	1	26	다크나이트 라이즈	(NULL)	0					
23	0	(NULL)	0	27	다크나이트 라이즈	(NULL)	0					
24	-6.344	(NULL)	1	29	다크나이트 라이즈	(NULL)	0					
25	-2.6973	(NULL)	1	30	다크나이트 라이즈	(NULL)	0					
26	0	(NULL)	0	31	다크나이트 라이즈	(NULL)	0					
27	0	(NULL)	0	33	다크나이트 라이즈	(NULL)	0					
28	0.8596000000000001	(NULL)	1	34	다크나이트 라이즈	(NULL)	0					
29	0.8303	(NULL)	1	35	다크나이트 라이즈	(NULL)	0					
30	0	(NULL)	0	36	다크나이트 라이즈	(NULL)	0					
31	0	(NULL)	0	37	다크나이트 라이즈	(NULL)	0					
32	0	(NULL)	0	38	다크나이트 라이즈	(NULL)	0					
33	0	(NULL)	0	39	다크나이트 라이즈	(NULL)	0					
34	21.9979	(NULL)	1	40	다크나이트 라이즈	(NULL)	0					
35	0	(NULL)	0	41	다크나이트 라이즈	(NULL)	0					
36	14.1372	(NULL)	1	42	다크나이트 라이즈	(NULL)	0					
37	0	(NULL)	0	43	다크나이트 라이즈	(NULL)	0					
38	-2.6973	(NULL)	1	44	다크나이트 라이즈	(NULL)	0					
39	0	(NULL)	0	45	다크나이트 라이즈	(NULL)	0					
40	0	(NULL)	0	46	다크나이트 라이즈	(NULL)	0					
41	8.8749	(NULL)	1	47	다크나이트 라이즈	(NULL)	0					
42	0	(NULL)	0	48	다크나이트 라이즈	(NULL)	0					
43	0	(NULL)	0	49	다크나이트 라이즈	(NULL)	0					
44	0.8303	(NULL)	1	50	다크나이트 라이즈	(NULL)	0					
45	0	(NULL)	0	51	다크나이트 라이즈	(NULL)	0					
46	0	(NULL)	0	52	다크나이트 라이즈	(NULL)	0					
47	11.1825	(NULL)	1	53	다크나이트 라이즈	(NULL)	0					

[그림 15] result, num\_result 테이블

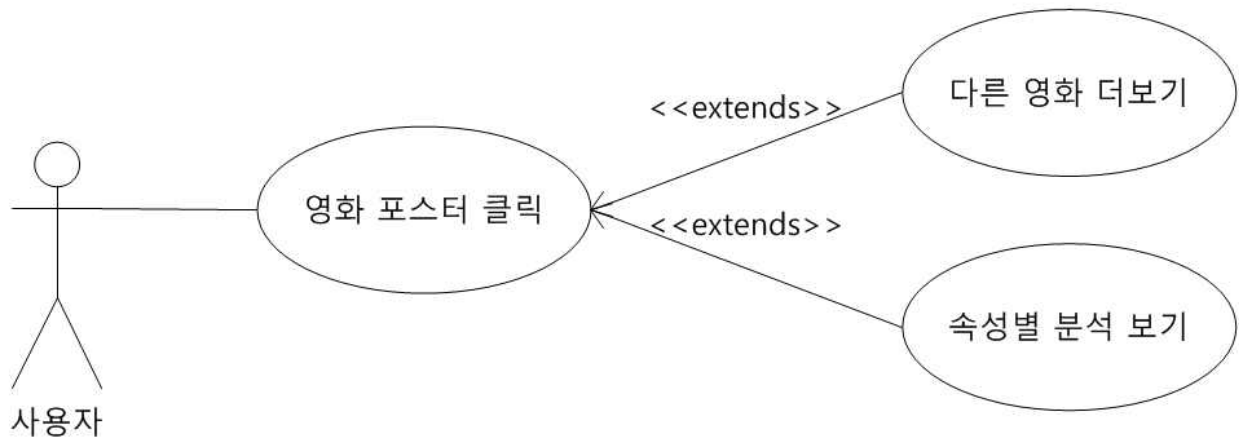
[그림 15]에는 result 테이블과 num\_result 테이블을 나타낸다. result 테이블은 미투데이 또는 트위터의 사용자 의견에 대한 선호도를 저장하고 선호도가 0인 것의 emotional\_flag는 0으로 설정하여 감정 문장이 아니라고 처리한다. 그리고 속성을 포함하는 경우 속성을 저장한다. num\_result 테이블은 각 영화별, 속성별로 긍정의 총합과 부정의 총합을 구하고 그 수치를 이용해 도표를 그리는 데 이용한다.

14	다크나이트 라이즈	괜찮	14
15	다크나이트 라이즈	더럽	3
16	다크나이트 라이즈	부럽	2
17	다크나이트 라이즈	반갑	2
18	다크나이트 라이즈	먹먹하	3
19	도둑들	멋지	20
20	도둑들	재밌	273
21	도둑들	귀엽	5
22	도둑들	좋	101
23	도둑들	괜찮	77

[그림 16] num\_adjective 테이블

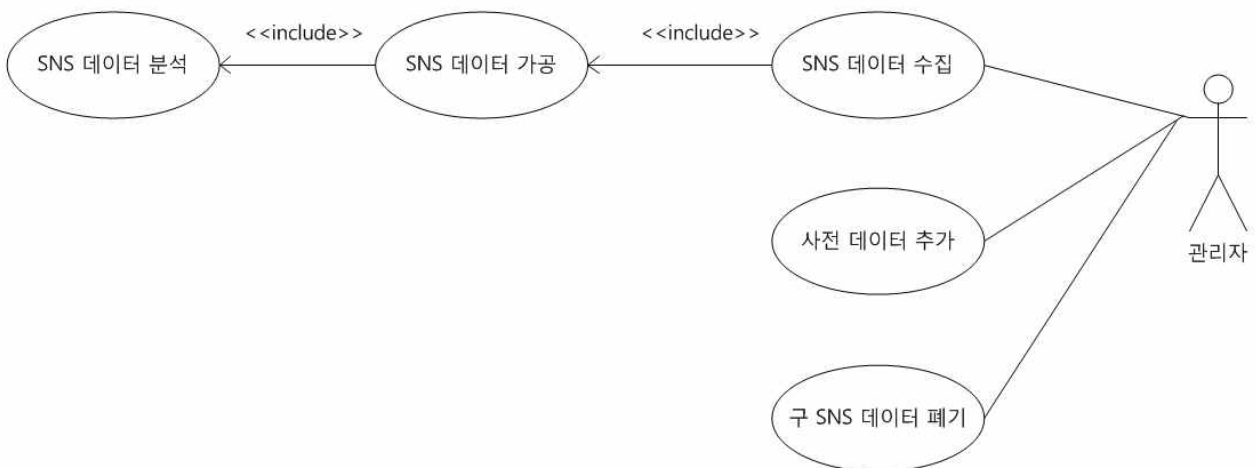
[그림 16]에서는 영화별로 텍스트데이터 내에 형용사가 나오는 빈도를 나타내는 num\_adjective 테이블이다.

### 3-4. Use Case Diagram



[그림 17] 사용자 Use Case Diagram

[그림 17]에서 볼 수 있듯이 사용자는 영화 포스터를 클릭하고 영화의 평판 분석 결과를 볼 수 있다. 영화 평판 분석 결과에서는 다른 영화 더 보기 기능과 영화의 속성별 분석 보기 기능이 포함되어 있다.



[그림 18] 관리자 Use Case Diagram

[그림 18]은 관리자의 Use Case Diagram이다. 관리자는 주기적으로 트위터와 미투데이의 SNS 데이터를 수집하여 데이터 가공 및 분석을 수행한다. 그리고 분석의 신뢰도를 높이기 위해 계속해서 감성 단어나 신조어를 사전에 입력한다. 또한 관리자는 더 이상 상영되지 않는 영화의 SNS 데이터를 폐기함으로써 스토리지 리소스를 최적화시킨다.



## 3-5. 테스트 설계

### 3-5-1. Unit Test

#### - 클라이언트

##### - WebPage 클래스

더미 서블릿 클래스가 더미 데이터를 돌려주도록 만들어 둔 다음 웹페이지에 접근해 올바른 내용이 뜨는지 확인한다.

더미 데이터는 다음과 같다.

```
<?xml version=W"1.0W" encoding=W"UTF-8W" ?><keyObjectList><keyObject name="형사"
thumbnail="http://www.movie.com/detective.jpg"/><keyObject name="블랙"
thumbnail="http://www.movie.com/black.jpg"/></keyObjectList>
```

더미 데이터에 의해 화면에 내용이 나타나면, 다시 더미 서블릿 클래스가 더미 데이터를 돌려주도록 만들어 둔 다음 리스트 컴포넌트의 항목을 클릭해 화면에 올바른 내용이 뜨는지 확인한다.

더미 데이터는 다음과 같다.

```
<?xml version=W"1.0W" encoding=W"UTF-8W" ?><analyzedOpinion><totalPreference
positive="827.690000" negative="-583.980000"/><opinion text="어벤져스 좋아했어거같다."
preference="1.130000"/><opinion text="오 어벤져스"
preference="-2.300000"/></analyzedOpinion>
```

##### - KeyObjectList 클래스

더미 클라이언트가 HTTP 연결로 KeyObjectList 클래스를 호출하게 한다. 처리 내용을 출력해 올바른 내용인지 확인한다.

##### - AnalyzedOpinion 클래스

더미 클라이언트가 HTTP 연결로 AnalyzedOpinion 클래스를 호출하게 한다. 처리 내용을 출력해 올바른 내용인지 확인한다.

#### - 띄어쓰기 모듈

더미데이터를 만들어서 띄어쓰기를 실행한다.

더미데이터의 일부 (예시)

<내 아내의모든 것, 이라는 제목은 다소 무게감있는데 비해 마지막까지도 그리무겁지 않게 즐길 수 있었다. 대부분의 이런영화가 그렇듯마지막엔 뭔가의미를 부여하곤 하지만... 전반적으로재밌게 즐길 수있는작품.>

#### - 문장분석 모듈

더미데이터의 분석을 실행한다.

더미데이터의 일부 (예시)

<내 아내의 모든 것, 이라는 제목은 다소 무게감 있는데 비해 마지막까지도 그리 무겁지 않게 즐길 수 있었다. 대부분의 이런 영화가 그렇듯 마지막엔 뭔가 의미를 부여하곤 하지만... 전반적으로 재밌게 즐길 수 있는 작품.>

#### - 수집 모듈

예상 된 수집데이터양과 실제 수집된 데이터의 양을 비교한다.

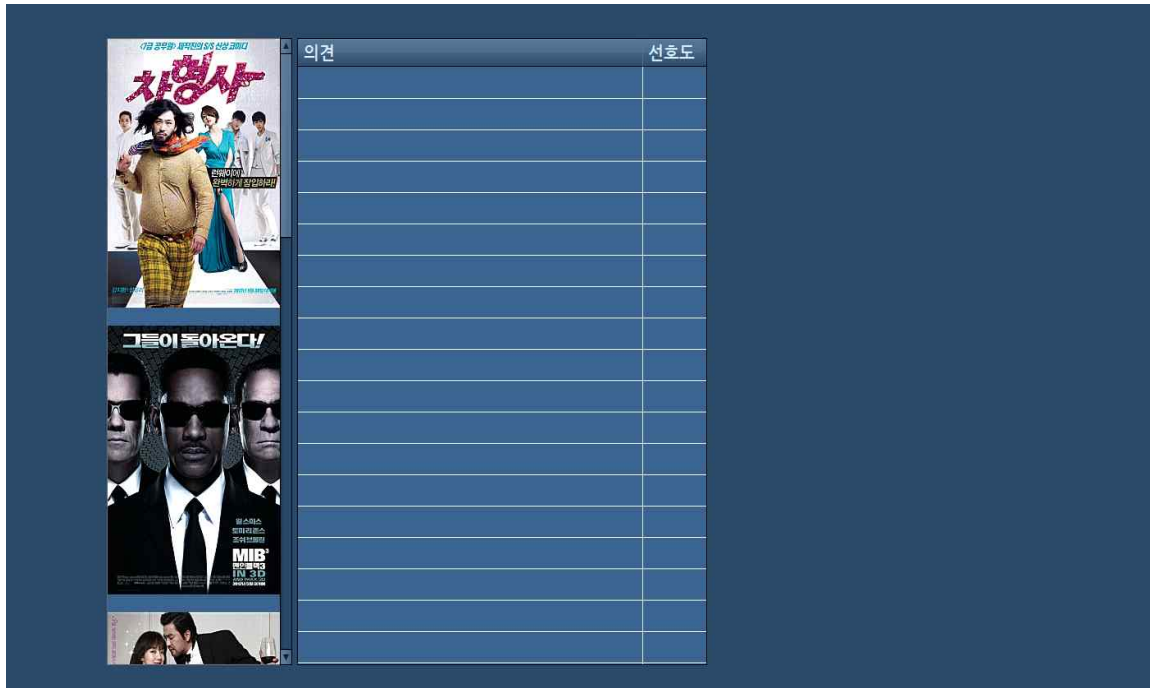
미투데이의 일주일 분량의 데이터를 한 번에 수집해본다.

트위터의 20page 분량의 데이터를 한 번에 수집해본다. (1page당 약 50개의 문장)

## 4. 최종 결과

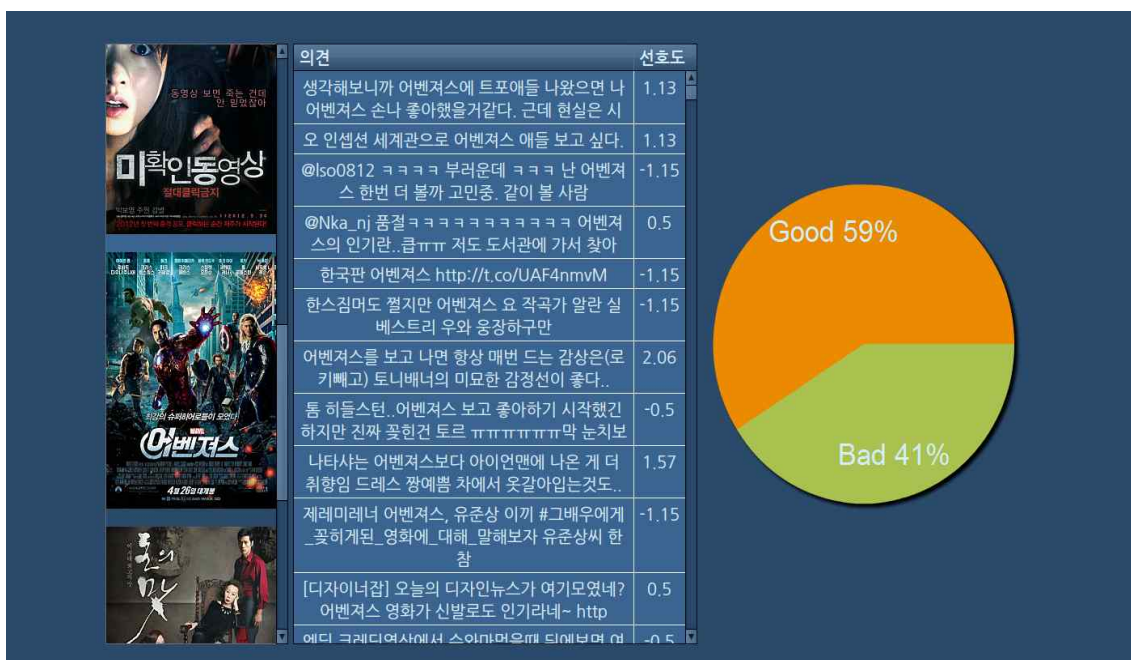
#### 4-1. 1학기

#### 4-1-1. UI



[그림 19] 최초 화면

사용자가 웹페이지에 접근하면 처음에 [그림 19]와 같은 화면에 나타난다. 화면 좌측에 Flex의 리스트 컴포넌트가 있고 그 안에 현재 분석 결과를 볼 수 있는 영화의 리스트가 들어있다. 스크롤바를 내려 분석 결과를 보고 싶은 영화를 클릭하면 영화에 대한 분석 결과를 볼 수 있다.



[그림 20] 결과 화면

영화를 클릭하면 [그림 20]처럼 분석 결과가 영화 리스트 오른쪽에 나타난다. 분석 결과의 상단에는 영화에 대한 전체 의견 중 긍정적인 감정 수치와 부정적인 감정 수치의 비율이 파이 그래프로 나타난다. 분석 결과 하단에는 전체 의견을 이루는 각 의견과 각 의견에서 계산되는 영화에 대한 선호도가 출력된다.

#### 4-1-2. 수집 모듈

13253	후궁 제왕의 첩을 보고...	2012-06-05T01:32:55+0900	후궁	1
13252	[후궁성형술,활기찬병원]활기찬병원에서 알려드리...	2012-06-04T17:26:18+0900	후궁	1
13251	후궁	2012-06-05T18:26:35+0900	후궁	1
13250	투어코리아 - [6월 개봉영화] '후궁: 제왕의 첩 6일 판...	2012-06-04T18:46:47+0900	후궁	1
13249	후궁:제왕의 첩	2012-06-05T14:26:07+0900	후궁	1
13248	[까칠이소식] '에로틱 사극'이라더니 이런 걸 숨겨 ...	2012-06-04T19:45:47+0900	후궁	1
13247	후궁 무대인사 정보	2012-06-05T12:58:57+0900	후궁	1
13246	&#39;후궁&#39;; 조여정 "마지막 정사신만 4일, 녹초...	2012-06-04T20:30:56+0900	후궁	1
13245	[후궁:제왕의 첩]대전도시철도공사와 함께하는 시사...	2012-06-05T02:19:55+0900	후궁	1
13244	오 후궁 개봉했네! 25:30 예약- ππ	2012-06-05T11:02:16+0900	후궁	1
13243	보고싶은 영화 후궁 롯데백화점 탈모	2012-06-05T13:41:09+0900	후궁	1
13242	'후궁' 임권택 감독과 최고 스타들의 만장일치 찬사'...	2012-06-05T10:15:22+0900	후궁	1
13241	잔인한 궁중권력다툼 속 살기 위해 처절히 변해야했...	2012-06-09T14:20:12+0900	후궁	1
13240	[볼스나인] 후궁 제왕의 첩!! 누적 관객수 100만 돌파...	2012-06-05T12:16:24+0900	후궁	1
13239	[ 후궁 , 제왕의 첩 ] - 후궁 시사회 후기	2012-06-05T18:13:25+0900	후궁	1
13238	보고 싶은 영화가 풍년이다. 간기남, 후궁,은교,돈의 맛.	2012-04-22T15:52:51+0900	후궁	1
13237	①후궁(+78) ②마이바흐(+240) ③미상민(+126) ④짝...	2012-05-17T13:00:01+0900	후궁	1
13236	방금영화보고왔어요.. 실망..ππ 너무 너무 아쉬움...	2012-06-08T23:08:38+0900	후궁	1
13235	2012년 5월 21일 후궁 시사회..	2012-06-05T23:04:10+0900	후궁	1
13234	<후궁: 제왕의 첩> 국내 최초 경희궁 초대형 쇼케이스...	2012-05-02T13:51:07+0900	후궁	1
13233	[롯데멤버스] <후궁,제왕의 첩>시사회 이벤트 350...	2012-05-03T09:00:55+0900	후궁	1
13232	<후궁:제왕의 첩>패션쇼 쇼케이스에 이웃분들 초대...	2012-05-04T16:14:13+0900	후궁	1
13231	[이벤트] 영화 "후궁:제왕의 첩" 제작보고회 & 패션쇼...	2012-05-03T15:01:03+0900	후궁	1
13230	[영화] 후궁 예고	2012-05-09T08:29:47+0900	후궁	1
13229	후궁: 제왕의 첩 패션쇼 쇼케이스 초대!!	2012-05-08T09:40:25+0900	후궁	1
13228	[발표] 영화 후궁 패션쇼 쇼케이스 오실분	2012-05-08T08:49:04+0900	후궁	1
13227	'후궁', 광기의 정사 담은 본 포스터 공개	2012-05-08T16:03:14+0900	후궁	1
13226	'후궁', 조여정 농염미 폭발..본 포스터 공개	2012-05-08T15:29:23+0900	후궁	1
13225	조여정이 또 사극에 나오네요. 후궁...	2012-05-09T13:58:21+0900	후궁	1
13224	<후궁: 제왕의 첩> 김동욱, 제왕의 광기로 물들다	2012-05-09T12:01:54+0900	후궁	1
13223	http://j.mp/Kc3s4j 영화 후궁; 제왕의 첩 에 대하여	2012-05-08T21:25:43+0900	후궁	1
13222	&#039;후궁&#039;; 조여정 농염미 폭발..본 포스터 ...	2012-05-08T16:30:22+0900	후궁	1
13221	[이벤트] 영화 " 후궁:제왕의 첩 " 제작보고회 & 패...	2012-05-08T21:08:27+0900	후궁	1
13220	후궁 제왕의 첩 예고편, 방자전까지	2012-05-10T20:00:18+0900	후궁	1
13219	[이벤트] 영화 " 후궁:제왕의 첩 " 제작보고회 & 패...	2012-05-08T22:33:26+0900	후궁	1
13218	목디스크치료사례★목디스크로 하지마비까지..lamin...	2012-05-10T10:40:13+0900	후궁	1
13217	①신세계 채용(+261) ②간루루(+495) ③이호준(+606...	2012-04-27T15:00:00+0900	후궁	1
13216	후궁 노출 줄거리	2012-05-12T09:58:36+0900	후궁	1
13215	<후궁: 제왕의 첩> 국내 최초 경희궁 초대형 쇼케이스...	2012-05-02T13:51:07+0900	후궁	1
13214	[롯데멤버스] <후궁,제왕의 첩>시사회 이벤트 350...	2012-05-03T09:00:55+0900	후궁	1
13213	<후궁:제왕의 첩>패션쇼 쇼케이스에 이웃분들 초대...	2012-05-04T16:14:13+0900	후궁	1
13212	[이벤트] 영화 "후궁:제왕의 첩" 제작보고회 & 패션쇼...	2012-05-03T15:01:03+0900	후궁	1
13211	[영화] 후궁 예고	2012-05-09T08:29:47+0900	후궁	1
13210	후궁: 제왕의 첩 패션쇼 쇼케이스 초대!!	2012-05-08T09:40:25+0900	후궁	1
13209	[발표] 영화 후궁 패션쇼 쇼케이스 오실분	2012-05-08T08:49:04+0900	후궁	1
13208	'후궁', 광기의 정사 담은 본 포스터 공개	2012-05-08T16:03:14+0900	후궁	1
13207	'후궁', 조여정 농염미 폭발..본 포스터 공개	2012-05-08T15:29:23+0900	후궁	1

[그림 21] 수집모듈 수집 결과



[그림 21]은 me2day에서 ‘후궁’이라는 영화에 대한 글을 수집모듈로 수집한 결과이다. 7개의 영화에 대해 미투 13253개, 멘션 10512개를 수집했다. 1주일간의 미투와, 트위터는 시간단위 접근을 처리하지 못하여 20page의 멘션을 수집했다. 미투는 후궁 : 1310개, 차형사 : 1967개, 내아내의모든 것 : 1481개, 맨인블랙3 : 1412개, 어벤져스 : 2772개, 돈의맛 : 1940개, 미확인동영상 : 1552개를 수집했으며, 멘션은 후궁 : 1780개, 차형사 : 1362개, 내아내의모든 것 : 1414개, 맨인블랙3 : 1737개, 어벤져스 : 1640개, 돈의맛 : 779개, 미확인동영상 : 1772개를 수집했다.

#### 4-1-3. 띄어쓰기 모듈

13253	10438	후궁 제왕의 첩을 보고...	후궁
13252	10437	[후궁성형술, 활기찬 병원]활기찬 병원에서 알려드...	후궁
13251	10436	후궁	후궁
13249	10435	후궁:제왕의 첩	후궁
13248	10434	[까칠이 소식] '에로틱 사극'이라더니 이런 걸 숨겨 ...	후궁
13247	10433	후궁 무대인 사 정보	후궁
13246	10432	&#39;후궁 &#39;; 조여 정 "마지막 정사신만 4일, 녹...	후궁
13245	10431	[후궁:제왕의 첩]대전도시 철도 공사와 함께하는 시...	후궁
13244	10430	오 후궁 개봉했네! 25:30 예약 - ㅠㅜ	후궁
13243	10429	보고싶은 영화 후궁 롯데백화점 탈모	후궁
13242	10428	'후궁' 임권택 감독과 최고 스타들의 만장 일치 찬사'...	후궁
13241	10427	잔인한 궁중권력 다툼 속 살기 위해 처절히 변해야 ...	후궁
13240	10426	[볼스나인] 후궁 제왕의 첩!! 누적 관객수 100만 돌...	후궁
13239	10425	[후궁, 제왕의 첩] - 후궁 시 사회 후기	후궁
13238	10424	보고 싶은 영화가 풍년이다. 간기남, 후궁, 은 교, 돈...	후궁
13237	10423	① 후궁(+78) ②마이 바흐(+240) ③이상민(+126) ④...	후궁
13236	10422	방금 영화보고 왔어요. . 실망. . ㅠㅜ 너무 너무 아쉬...	후궁
13235	10421	2012년 5월 21일 후궁 시 사회. .	후궁
13234	10420	< 후궁: 제왕의 첩> 국내 최초 경희궁 초대형 쇼케이...	후궁
13233	10419	[롯데멤버스] <후궁, 제왕의 첩>시 사회 이벤트 3...	후궁
13232	10418	< 후궁:제왕의 첩>패션쇼 쇼케이스에 이웃분들 초대...	후궁
13231	10417	[이벤트] 영화 "후궁:제왕의 첩" 제작 보고 회 & 패...	후궁
13230	10416	[영화] 후궁 예고	후궁
13229	10415	후궁: 제왕의 첩 패션쇼 쇼케이스 초대!!	후궁
13228	10414	[발표] 영화 후궁 패션쇼 쇼케이스 오실분	후궁
13227	10413	'후궁', 광기의 정사 담은 본 포스터 공개	후궁
13225	10412	조여정이 또 사극에 나오네요. 후궁...	후궁
13224	10411	< 후궁: 제왕의 첩> 김동욱, 제왕의 광기로 물들다!	후궁
13223	10410	http://j. mp/Kc3s4j 영화 후궁; 제왕의 첩 에 대하여	후궁
13222	10409	&#039;후궁 &#039;;, 조여 정 농염미 폭발. . 본 포스...	후궁
13221	10408	[이벤트] 영화 " 후궁:제왕의 첩 " 제작 보고 회 & 패...	후궁
13220	10407	후궁 제왕의 첩 예고 편, 방자전까지	후궁
13219	10406	[이벤트] 영화 " 후궁:제왕의 첩 " 제작 보고 회 & 패...	후궁
13218	10405	목디스크치료 사례★목디스크로 하지 마비까지. . la...	후궁
13217	10404	①신세계 채용(+261) ②간루루(+495) ③이호준(+606...	후궁
13216	10403	후궁 노출 줄거리	후궁
13215	10402	< 후궁: 제왕의 첩> 국내 최초 경희궁 초대형 쇼케이...	후궁
13214	10401	[롯데멤버스] <후궁, 제왕의 첩>시 사회 이벤트 3...	후궁
13213	10400	< 후궁:제왕의 첩>패션쇼 쇼케이스에 이웃분들 초대...	후궁
13212	10399	[이벤트] 영화 "후궁:제왕의 첩" 제작 보고 회 & 패...	후궁
13211	10398	[영화] 후궁 예고	후궁
13210	10397	후궁: 제왕의 첩 패션쇼 쇼케이스 초대!!	후궁
13209	10396	[발표] 영화 후궁 패션쇼 쇼케이스 오실분	후궁
13208	10395	'후궁', 광기의 정사 담은 본 포스터 공개	후궁
13206	10394	조여정이 또 사극에 나오네요. 후궁...	후궁
13205	10393	< 후궁: 제왕의 첩> 김동욱, 제왕의 광기로 물들다!	후궁
13204	10392	http://j. mp/Kc3s4j 영화 후궁; 제왕의 첩 에 대하여	후궁

[그림 22] 띄어쓰기 모듈을 이용한 띄어쓰기 결과

[그림 22]는 띄어쓰기 모듈을 이용한 띄어쓰기 결과로서, [그림 21]의 수집한 결과를 띄어쓰기 해서 새로 DB에 저장한 결과이다.

#### 4-1-4. 분석 모듈

posnav	twitter_id	emotional_flag	me2day_id	movie_name
2.66	(NULL)	1	1332	어벤져스
1.07	(NULL)	1	1333	어벤져스
0	(NULL)	0	1329	어벤져스
-0.5	(NULL)	1	1330	어벤져스
0	(NULL)	0	1331	어벤져스
0	(NULL)	0	1328	어벤져스
-0.5	(NULL)	1	1326	어벤져스
0	(NULL)	0	1327	어벤져스
0	(NULL)	0	1323	어벤져스
0	(NULL)	0	1324	어벤져스
-1.15	(NULL)	1	1325	어벤져스
0	(NULL)	0	1319	어벤져스
0	(NULL)	0	1320	어벤져스
0	(NULL)	0	1321	어벤져스
0	(NULL)	0	1322	어벤져스
0	(NULL)	0	1318	어벤져스
1.57	(NULL)	1	1314	어벤져스
0	(NULL)	0	1315	어벤져스
0	(NULL)	0	1316	어벤져스
0	(NULL)	0	1317	어벤져스
0	(NULL)	0	1312	어벤져스
0	(NULL)	0	1313	어벤져스
0	(NULL)	0	1308	어벤져스
0	(NULL)	0	1309	어벤져스
-1.65	(NULL)	1	1310	어벤져스
0	(NULL)	0	1311	어벤져스
0	(NULL)	0	1307	어벤져스
0	(NULL)	0	1306	어벤져스
-1.15	(NULL)	1	1296	어벤져스
0	(NULL)	0	1305	어벤져스
0	(NULL)	0	1297	어벤져스
0	(NULL)	0	1298	어벤져스
0	(NULL)	0	1299	어벤져스
2.46	(NULL)	1	1300	어벤져스
-0.5	(NULL)	1	1302	어벤져스
0	(NULL)	0	1303	어벤져스
0	(NULL)	0	1289	어벤져스
-0.5	(NULL)	1	1290	어벤져스
0	(NULL)	0	1291	어벤져스
-1.15	(NULL)	1	1292	어벤져스
0	(NULL)	0	1293	어벤져스
3.14	(NULL)	1	1294	어벤져스
0	(NULL)	0	1295	어벤져스
0	(NULL)	0	1288	어벤져스
0	(NULL)	0	1285	어벤져스
0	(NULL)	0	1286	어벤져스
1.57	(NULL)	1	1287	어벤져스

[그림 23] 분석모듈 분석 결과1

[그림 23]은 분석모듈을 통한 분석 결과를 보여주는 것이며, 각각의 미투와 멘션에 대해 감성문장 인지 아닌지 분석하고 감성문장이라면 감성 수치를 부여한 결과이다.

id	pos_sum	nav_sum	movie_name
1	800.340000000005	-302.31	차형사
2	455.469999999999	-238.21	맨인블랙3
3	(NULL)	(NULL)	내아내의모든것
4	326.999999999999	-331.12	미확인동영상
5	827.69	-583.979999999996	어벤져스
6	556.279999999999	-676.179999999999	돈의맛
7	350.479999999999	-182.22	후궁

[그림 24] 분석모듈 분석 결과 2

[그림 24]는 각각의 영화에 대해 긍정값과 부정값의 총점을 계산해서 저장한 결과이다. 이러한 총점은 결과화면의 도표를 출력하는데 이용된다.

#### 4-1-5. 감성단어 사전

id	text	real	friendly	posnav	active
1	가쁜	3.37	3.26	4.51	3.8
2	가련	3.92	3.87	3.11	3.17
3	가소롭	3.88	3.26	2.48	3.91
4	가엎	4.54	3.87	3	3.33
5	가증	4.1	4.67	1.88	5.08
6	가책	3.45	5.11	2.55	4.21
7	갈등	4.19	4.47	2.59	4.82
8	감개	3.78	4.13	4.96	4.24
9	감개무량	4.29	3.9	4.97	4.83
10	감격	5.07	5.46	5.43	5.19
11	감동	5.24	5.5	5.45	4.47
12	감명	4.06	4.24	5.28	4.01
13	감미	4.93	4.88	5.54	3.93
14	감복	3.46	3.12	4.63	4.59
15	감탄	4.79	5.18	5.11	5.03
16	감흥	3.48	3.25	4.59	4.21
17	감회	3.95	4.05	4.67	3.98
18	개탄	3.35	3.08	2.33	5.18
19	거부감	4.26	4.82	1.98	4.54
20	거북	3.92	4.66	2.58	4.14

[그림 25] 감성단어 사전 샘플

[그림 25]는 감성문장 분류 모듈에 사용하는 감성단어들을 나타낸다. 표준어 감성단어 424개의 패볼패와 원형성과 친숙도와 활성도가 저장되어 있다. 전체 감성단어 사전은 첨부D에 첨부되어 있다.

#### 4-1-6. Mapping Table

id	dic_id	text
119	63	간지나다
140	63	감사
142	63	감좌
98	128	감지덕지
77	4	갑갑
154	63	강추
137	63	고상
136	63	고상하다
135	63	고상한
91	128	공감
106	128	괜찮다

[그림 26] Mapping Table 샘플

[그림 26]은 Mapping Table에 있는 단어의 샘플이다. Mapping Table에는 표준어에 포함되어 있지 않고 영화에 대한 감성을 표현하는 감성단어가 저장되어 있다. Mapping Table에 들어간 전체 단어는 '첨부B.설문조사'에 있다.



## 4-2. 2학기

### 4-2-1. 웹페이지



[그림 27] 최초 화면

웹페이지에 접속하면 위 [그림 27]과 같은 화면을 볼 수 있다. 화면의 정중앙에는 프로젝트 로고가 삽입되고 있고 그 아래로 최근 상영중인 영화 포스터들이 가로로 배열되어 있다. 영화 포스터는 총 10개로 이루어져 있고 사용자는 포스터 양쪽에 있는 화살표를 통해 다른 영화 포스터를 선택할 수 있다. 사용자가 원하는 포스터를 선택하면 결과 화면 웹페이지로 이동한다.



[그림 28] 영화 분석 화면

[그림 28]은 위에서 언급한 초기 화면에서 사용자가 영화 포스터를 선택했을 때 나타나는 결과 화면이다. 본 페이지는 사용자가 선택한 영화의 긍부정 분석 결과를 출력하는데 목적이 있다. 화면 좌측에는 사용자가 선택한 영화의 포스터가 나타나고 그 아래로는 영화에 대한 긍정 비율이 높은 순으로 나머지 영화들의 순위가 매겨져 있다. 화면의 중앙에는 분석 결과가 나타나는데 먼저 상단에는 원 그래프를 통해 영화의 전체 긍부정 비율과 막대 그래프를 통한 영화의 속성 긍부정 비율이 나타난다. 그래프 아래에는 태그 클라우드가 있어서 SNS 문장에 포함되어있는 형용사의 개수를 폰트 크기와 글자 색으로 시각화한다. 태그 클라우드 아래에는 해당 영화를 언급한 SNS 사용자의 의견을 시간순으로 정렬하여 나열함으로서 분석 결과에 대한 사용자의 이해를 돕는다.

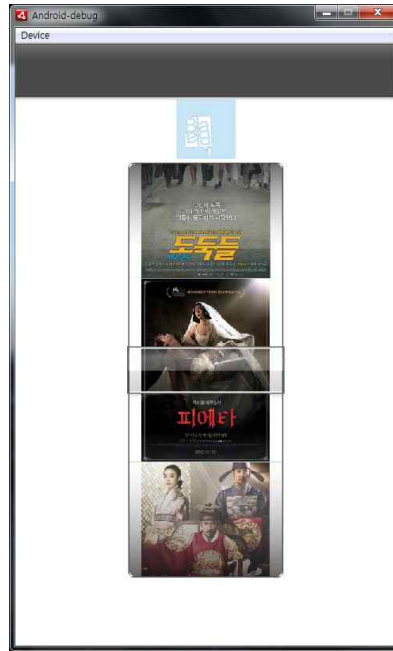


[그림 29] 형용사별 영화 순위 화면

[그림 29]는 결과 화면 오른쪽에 있는 형용사 리스트의 형용사 중 하나를 클릭했을 때 나타나는 화면이다. 형용사를 클릭하면 SNS에 올라온 각 영화를 언급한 문장들에서 해당 형용사가 얼마나 등장하였는지에 대한 순위가 나온다.

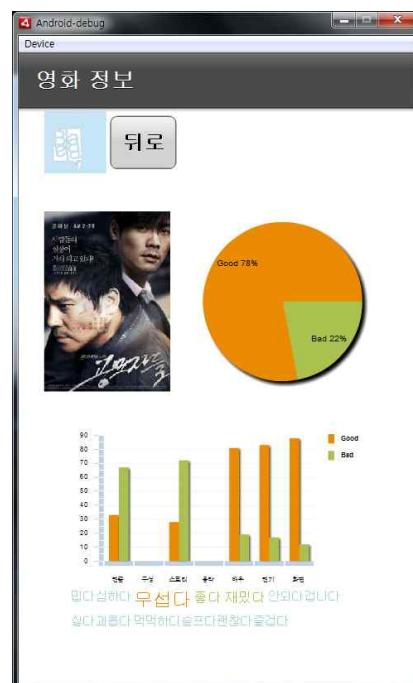


## 4-2-2. 안드로이드 애플리케이션



[그림 30] 최초 화면

안드로이드 애플리케이션을 실행하면 위 [그림 30]과 같은 화면을 볼 수 있다. 화면 가운데의 리스트를 움직여 원하는 영화를 터치하면 원하는 영화의 평판 분석 결과를 볼 수 있다.



[그림 31] 영화 분석 화면

[그림 31]은 원하는 영화를 선택했을 때 나타나는 화면이다. 파이 그래프를 통해 영화에 대한 전체적인 선호도를 알 수 있고 아래의 막대 그래프를 통해 영화의 각 속성에 대한 선호도를 알 수 있다. 그리고 막대 그래프 아래에는 태그 클라우드가 있어서 SNS 문장에 포함되어있는 형용사의 개수를 폰트 크기와 글자 색으로 시각화한다.

### 4-2-3. 속성 분석 모듈

id	pos_sum	nav_sum	movie_name	attribute
7	5240.976	-803.732000000001	다크나이트 라이즈	(NULL)
17	63.7913	(NULL)	다크나이트 라이즈	연출
27	(NULL)	-6.7823	다크나이트 라이즈	구성
37	78.871	-19.2348	다크나이트 라이즈	스토리
47	26.7542	-10.7369	다크나이트 라이즈	음악
57	83.1429	-2.6973	다크나이트 라이즈	배우
67	18.0352	(NULL)	다크나이트 라이즈	연기
77	(NULL)	(NULL)	다크나이트 라이즈	화면
1	8281.002700000006	-1502.3924	도둑들	(NULL)
11	46.4253	-4.9909	도둑들	연출
21	41.8354	-5.3946	도둑들	구성
31	250.7365	-5.916	도둑들	스토리
41	15.2174	(NULL)	도둑들	음악
51	400.907	-13.6598	도둑들	배우
61	225.9033	(NULL)	도둑들	연기
71	(NULL)	(NULL)	도둑들	화면

[그림 32] 속성별 분석 결과

위 [그림 32]는 속성별 분석 결과 정보를 나타낸다. 각 각의 영화에 대해 세부 항목으로 다양한 속성들이 삽입되어 있는 것을 확인할 수 있다. 'attribute'가 'NULL'인 것은 영화에 대한 전체 긍정 분석 결과를 나타내고, 그렇지 않은 것은 각 각의 속성에 대한 긍정 분석 결과를 나타낸다. 분석 결과 웹페이지 및 안드로이드 애플리케이션에서는 이 테이블의 'pos\_sum'과 'nav\_sum' 값을 백분율로 변환하여 사용하게 된다.

### 4-2-4. 형태소 분석 모듈

id	movie_name	adjective	num_adjective
16	도둑들	겁나	17
33	무서운이야기	겁나	6
47	연가시	겁나	9
58	새미익어드벤처2	겁나	2
71	락오브에이지	겁나	9
75	다크나이트 라이즈	겁나	3
98	말리	겁나	3
122	매직마이크	겁나	3
59	새미익어드벤처2	고맙	3
5	도둑들	괜찮	77
20	무서운이야기	괜찮	17
42	연가시	괜찮	25
52	새미익어드벤처2	괜찮	4
67	락오브에이지	괜찮	4
88	다크나이트 라이즈	괜찮	14
103	말리	괜찮	7
117	매직마이크	괜찮	6

[그림 33] 형태소 분석 결과

긍정 분석 결과와 함께 웹페이지 및 안드로이드 애플리케이션에 제공될 형태소 분석 결과는 위 [그림 33]과 같은 구조로 이루어져 있다. 각 영화의 SNS 의견에서 사용된 감성 단어의 빈도를 저장하고 이 정보를 활용하여 감성 단어에 따른 영화 순위 또는 영화에 따른 감성 단어 순위로 사용자에게 도움이 될 수 있는 정보를 제공한다.

#### 4-2-5. 속성 단어 매핑 테이블

id	attribute	id	attr_id	value
1	연출	1	1	표현력
2	구성	2	2	구도
10	스토리	4	10	줄거리
6	음악	5	10	내용
7	배우	6	6	음향
8	연기	7	6	BGM
9	화면	8	6	브금

[그림 34] 속성 테이블과 속성 단어 매핑 테이블

[그림 34]의 좌측 테이블은 속성 테이블이다. 속성 테이블은 미리 속성 값이 지정되어 있는데 SNS 사용자들은 하나의 속성을 다양한 단어로 표현하는 경우가 많기 때문에(예. 스토리, 줄거리) 우측 테이블과 같이 매핑 테이블을 만들어 속성 단어를 매핑시키는 방법으로 이같은 문제를 해결하였다.

#### 4-2-6. 자카드 계수를 활용한 감성 단어 매핑

본 프로젝트에서는 SNS에 새롭게 나타나는 줄임말이나 은어와 같은 감성 단어를 매핑 테이블에 추가하기 위해 '자카드 계수(Jaccard coefficient)'를 사용한다. 자카드 계수의 측정 공식은 아래 [수식 2]와 같다.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

[수식 2] 자카드 계수 측정 공식

예를 들어, 두 개의 감성 단어 '재미있다'와 '재미있다'를 통해 자카드 계수를 측정하면 세부 원리는 아래 [표 5]와 같다.

구분	재미있다	재미있다
집합 1	ㅈ, ㅊ	ㅈ, ㅊ
집합 2	ㅊ, ㅊ	ㅊ, ㅊ
집합 3	ㅊ, ㅣ	ㅊ, ㅣ
집합 4	ㅣ, ㅇ	ㅣ, ㅇ
집합 5	ㅇ, ㅣ	ㅇ, ㅣ
집합 6	ㅣ, ㅅ	ㅣ, ㅅ
집합 7	ㅅ, ㅈ	ㅅ, ㅈ
집합 8	ㅈ, ㅈ	ㅈ, ㅈ

[표 5] 자카드 계수 측정 원리

먼저 단어의 음절을 각각 초성과 중성 그리고 종성으로 나눈 후 위와 같이 두 개씩 집합을 구성한다. 그리고 위의 집합을 토대로 합집합과 교집합의 개수를 세는데 '재미있다'와 '재밌다'의 경우 교집합의 개수는 6개, 합집합의 개수는 10개라는 것을 알 수 있다. 이를 [수식 1]에 대입하면 아래와 같다.

$$J(\text{재미있다}, \text{재밌다}) = \frac{6}{10} = 0.6$$

[수식 3] '재미있다'와 '재밌다'의 자카드 계수

#### 4-2-7. URL 제거를 통한 광고 제거 모듈 성능 향상

기존의 광고 제거 모듈은 내용이 완전히 일치하는 문장이 2개 이상 있을 경우, 이를 광고로 판단하고 분석 대상에서 제외시킨다. 하지만 최근 SNS에 나타나는 광고는 지능적인 방법을 통해 이와 같은 광고 제거 기법을 회피한다.

내용	해시
레지던트5 이벤트! <a href="http://t.co/syohFK9i">http://t.co/syohFK9i</a>	92812355
레지던트5 이벤트! <a href="http://t.co/X4mgS0ew">http://t.co/X4mgS0ew</a>	51821852
레지던트5 이벤트! <a href="http://t.co/9MNY9Yer">http://t.co/9MNY9Yer</a>	-5215855
레지던트5 이벤트! <a href="http://t.co/vh6YxP1v">http://t.co/vh6YxP1v</a>	21263812
레지던트5 이벤트! <a href="http://t.co/mE6F3Ega">http://t.co/mE6F3Ega</a>	-1821251
레지던트5 이벤트! <a href="http://t.co/IQILPTyl">http://t.co/IQILPTyl</a>	-2288325

[표 6] 내용은 같지만 해시가 다른 광고 문서 예시

위의 [표 6]을 살펴보면, 내용은 같지만 마지막에 포함되어 있는 URL 주소가 다르기 때문에 전혀 다른 해시값이 생성되는 것을 확인할 수 있다. 이와 같이 최근 SNS상의 광고는 서로 다른 URL 주소를 포함하게 하여 광고 문장들이 서로 같지 않게 인식되도록 광고 제거 모듈을 회피하고 있다. 그래서 본 프로젝트에서는 이런 문제점을 간파하고 광고 제거 모듈의 성능을 향상시키기 위해 광고를 제거하기 전에 미리 문장에 포함된 URL을 제거한 후 내용을 비교하도록 모듈의 성능을 개선하였다. 개선 결과 임의로 텍스트 데이터 100개를 추출했을 때 전에는 광고의 비율이 14%였지만, 현재는 4%로 크게 감소하였다는 것을 확인할 수 있었다.

## 5. 평가 항목 및 방법

본 프로젝트는 극성의 분류와 강도를 구하기 위해 앞에서 사용했던 [수식 5]를 변형한 20가지 수식을 사용해 각 수식의 성능을 실험해 보았다. [수식 4]에서 집합  $W(d)$ 는 문서  $d$ 에서 추출된 형태소의 집합  $A(d)$ 와 감성 단어 집합  $E$ 의 교집합이다. 그리고  $word_{d,k}$ 는 집합  $W(d)$ 의 원소이며,  $|W(d)|$ 는 문서  $d$ 에 포함된 감성 단어의 수이다.  $posne()$ 는 단어의 쾌-불쾌 수치를 나타내는 함수,  $activ()$ 는 단어의 활성화 수치를 나타내는 함수,  $proto()$ 는 단어의 원형성 수치를 나타내는 함수,  $frien()$ 는 단어의 친숙성 수치를 나타내는 함수,  $pref(d)$ 는 문서  $d$ 의 선호도를 나타내는 함수라고 했을 때, [수식 5]는 쾌-불쾌 수치에서 3.81을 뺀 값이 양수이면 1을 사용하고 음수이면 -1을 사용한 식이다. [수식 6]은 [수식 5]에서 사용한 1 또는 -1에 활성화 수치를 곱한 값을 사용한 식이다. [수식 7]은 쾌-불쾌 수치에서 3.81을 뺀 값만 사용한 식이다. [수식 8]은 쾌-불쾌 수치에서 3.81을 뺀 값에 활성화 수치를 곱한 값을 사용한 식이다. [수식 9]는 [수식 7]에서 사용한 값과 [수식 6]에서 사용한 값을 더한 값을 사용한 식이다. [수식 5], [수식 6], [수식 7], [수식 8], [수식 9]와 이 다섯 가지 기본 수식에 원형성 수치를 반영하기 위해  $\frac{proto()}{7}$ 를 곱한 수치를 사용한 수식 다섯 가지, 기본 수식에 친숙성 수치를 반영하기 위해  $\frac{frien()}{7}$ 을 곱한 수치를 사용한 수식 다섯 가지, 기본 수식에 원형성 수치와 친숙성 수치를 모두 반영하기 위해  $\frac{proto()}{7} \times \frac{frien()}{7}$ 을 곱한 수치를 사용한 수식 다섯 가지를 사용했다. 원형성 수치는 단어가 감정 단어의 좋은 예일수록 수치를 많이 반영하기 위해 사용했고, 친숙성 수치는 [9]의 실험 결과 원형성 수치와 친숙성 수치가 높은 정적인 상관을 보였기 때문에 사용했다.

$$word_{d,k} \in W(d) = A(d) \cap E$$

수식 4 감성 단어 추출

$$pref(d) = \sum_{k=1}^{|W(d)|} \left( \frac{|posneg(word_{d,k}) - \alpha|}{posneg(word_{d,k}) - \alpha} \right)$$

수식 5 문서의 극성 분류 및 극성의  
강도 계산식 ( $\alpha = 3.81$ )

$$pref(d) = \sum_{k=1}^{|W(d)|} \left( \frac{|posneg(word_{d,k}) - \alpha|}{posneg(word_{d,k}) - \alpha} \times active(word_{d,k}) \right)$$

수식 6 문서의 극성 분류 및 극성의  
강도 계산식 ( $\alpha = 3.81$ )

$$pref(d) = \sum_{k=1}^{|W(d)|} (posneg(word_{d,k}) - \alpha)$$

수식 7 문서의 극성 분류 및 극성의  
강도 계산식 ( $\alpha = 3.81$ )

$$pref(d) = \sum_{k=1}^{|W(d)|} ((posneg(word_{d,k}) - \alpha) \times active(word_{d,k}))$$

수식 8 문서의 극성 분류 및 극성의  
강도 계산식 ( $\alpha = 3.81$ )

$$pref(d) = \sum_{k=1}^{|W(d)|} (((posneg(word_{d,k}) - \alpha) + \frac{|posneg(word_{d,k}) - \alpha|}{posneg(word_{d,k}) - \alpha}) \times active(word_{d,k}))$$

수식 9 문서의 극성 분류 및 극성의  
강도 계산식 ( $\alpha = 3.81$ )

## 5-1. 감성단어인식률

감성단어인식률이란, 문장 분석 모듈을 통하여 특정 문장이 감성 단어를 포함하는 문장인지 포함하지 않는 문장인지를 얼마나 잘 분류하는가를 측정하기 위한 기준으로 계산식은 아래 [수식 9]와 같다.

$$\text{감성단어인식률} = \frac{\text{올바르게 구분한 데이터 개수}}{\text{전체 데이터 개수}}$$

[수식 9] 감성단어인식률 계산식

모든 데이터에 대한 테스트는 상당한 시간적 비용을 소요하므로, 전체 데이터 중 임의의 샘플 데이터 100개를 추출하여 이에 대한 평가를 실시한다. 평가는 아래의 [그림 35]와 같은 방법으로 테스트가 직접 문장 분석 모듈의 결과와 실제 데이터를 비교한다.

Primary Key	테스터에 의한 분류	시스템에 의한 분류	일치 여부
1011	감성 문장	감성 문장	○
1012	비감성 문장	비감성 문장	○
1013	감성 문장	감성 문장	○
1014	비감성 문장	비감성 문장	○
1015	비감성 문장	감성 문장	×
1016	감성 문장	감성 문장	○
1017	비감성 문장	비감성 문장	○
1018	감성 문장	비감성 문장	×
...	...	...	...

[그림 35] 감성단어인식률 평가 예제

또한 매핑 테이블 사용 여부에 따른 감성단어인식률을 측정하기 위해 매핑 테이블을 사용하지 않았을 때의 정확도, 매핑 테이블에 설문조사에 포함된 단어만 입력하고 사용했을 때의 정확도, 매핑 테이블에 자카드 계수 0.6 이상인 단어 90개를 추가하고 사용했을 때의 정확도를 측정해 본다.

## 5-2. 극성분류정확도

극성분류정확도란, 문장 분석 모듈을 통해 감성 문장으로 분류된 문장들을 대상으로, 극성(긍정 또는 부정)을 얼마나 잘 분류했는가를 평가하기 위한 기준으로 계산식은 [수식 10]과 같다.

$$\text{극성분류정확도} = \frac{\text{올바른 극성을 가지는 감성 문장의 개수}}{\text{전체 감성 문장의 개수}}$$

[수식 10] 극성분류정확도 계산식

모든 데이터에 대한 테스트는 상당한 시간적 비용을 소요하므로, 전체 데이터 중 임의의 샘플 데이터 100개를 추출하여 이에 대한 평가를 실시한다. 평가는 아래의 [그림 36]과 같은 방법으로 테스트가 직접 문장 분석 모듈의 결과와 실제 데이터를 비교한다.

ID	테스터에 의한 분류	시스템에 의한 분류	일치 여부
1	긍정	긍정	○
2	부정	중립	×
3	긍정	긍정	○
4	중립	부정	×
5	중립	부정	×
6	긍정	긍정	○
7	긍정	긍정	○
8	부정	중립	×
...	...	...	...

[그림 36] 극성분류정확도 평가 예제

극성분류정확도 역시 감성단어인식률과 마찬가지로 매핑 테이블 사용 여부에 따른 극성분류정확도 정확도를 측정하기 위해 매핑 테이블을 사용하지 않았을 때의 정확도, 매핑 테이블에 설문조사에 포함된 단어만 입력하고 사용했을 때의 정확도, 매핑 테이블에 자카드 계수 0.6 이상인 단어 90개를 추가하고 사용했을 때의 정확도를 측정해 보았다.

## 5-3. 평가 결과

### 5-3-1. 감성단어인식률

평가 결과 매핑 테이블을 사용하지 않았을 때는 66%, 설문조사에 포함된 단어만 입력하고 매핑 테이블을 사용했을 때는 68%, 자카드 계수 0.6 이상인 단어 90개를 추가하고 매핑 테이블을 사용했을 때는 65%의 정확도를 보였다.

### 5-3-2. 극성분류정확도

측정 결과는 표 7과 표 8에 나타났다. [표 7]에서 [수식 5]는 1-1, [수식 5]에  $\frac{proto()}{7}$ 를 곱한 수치를 사용한 수식은 1-2,  $\frac{frien()}{7}$ 을 곱한 수치를 사용한 수식은 1-3,  $\frac{proto()}{7} \times \frac{frien()}{7}$ 을 곱한 수치를 사용한 수식은 1-4로 표기하고 [수식 6], [수식 7], [수식 8], [수식 9]에 대해서도 마찬가지로 표기하였다. 그리고 매핑 테이블을 사용하지 않았을 때의 정확도는 ‘사용 전’, 매핑 테이블에 설문조사에 포함된 단어만 입력하고 사용했을 때의 정확도는 ‘사용 후’, 매핑 테이블에 자카드 계수 0.6 이상인 단어 90개를 추가하고 사용했을 때의 정확도는 ‘추가 후’라고 표기한 열에 표기하였다. [표 8]은 감성 포함 여부 확인 정확도가 100%라고 가정했을 때의 정확도이다. 감성 포함 여부 확인 과정에서 올바르게 분류된 데이터만을 대상으로 극성 분류 정확도를 측정했다. 극성 분류 예는 [표 9]와 같다.

수식	사용 전	사용 후	추가 후
1-1	48	51	44
1-2	48	51	51
1-3	48	51	51
1-4	48	51	52
2-1	48	51	50
2-2	48	51	51
2-3	48	51	50
2-4	48	51	52
3-1	48	50	50
3-2	48	50	51
3-3	48	50	51
3-4	48	50	51
4-1	48	50	51
4-2	48	50	51
4-3	48	50	51
4-4	48	50	51
5-1	48	51	52
5-2	48	51	52
5-3	48	51	51
5-4	48	51	52

[표 7] 각 수식과 매핑 테이블 사용 여부에 따른 극성 분류 정확도 (단위 : %)



수식	사용 전	사용 후	추가 후
1-1	73	75	68
1-2	73	75	78
1-3	73	75	78
1-4	73	75	80
2-1	73	75	77
2-2	73	75	78
2-3	73	75	77
2-4	73	75	80
3-1	73	74	77
3-2	73	74	78
3-3	73	74	78
3-4	73	74	78
4-1	73	74	78
4-2	73	74	78
4-3	73	74	78
4-4	73	74	78
5-1	73	75	80
5-2	73	75	80
5-3	73	75	78
5-4	73	75	80

[표 8] 감성 포함 여부 확인 정확도가 100%일 때의 극성 분류 정확도 (단위 : %)

문서	시스템 분류	실제 분류
올 광해 재밌다	긍정	긍정
익스펜더블2 재미있다 ㅋㅋ 눈이 즐거워집니다 ㅋㅋ	긍정	긍정
음 늑대아이 또봐도 재밌다	긍정	긍정
@eunny_ 본레거시가 흥행참패를했지.. 손님 디게 없드라 ^_ㅠ 사람 정말 많을줄 알았는데..	부정	부정

[표 9] 극성 분류 예

## 6. 기대 효과 및 활용 방안

### 6-1. 기술적 측면

현재 임의의 주제에 대해 간단한 평판 분석을 수행하는 기술은 존재하지만, SNS상에서 사용자들이 주고받는 일상 대화를 표본으로 영화의 세부 속성에 대해 평판 분석을 수행하는 기술은 아직까지 활성화된 것이 없다. 이런 이유로 본 프로젝트는 위와 같은 기술을 개발 및 개선하여 영화라는 제한된 범위에서 벗어나 제품이나 인물 등 위 기술이 적용될 수 있는 다양한 범위의 평판 분석으로 확장될 수 있다.

### 6-2. 경제적, 산업적 측면

개인의 측면에서는 새로 개봉한 영화에 대한 정보를 얻기 위해 주로 개인 블로그나 포털 사이트의 영화 상품평을 읽는 것에 많은 시간을 투자한다. 하지만 본 프로젝트의 평판 분석 기술을 사용하면 정보를 검색하는 시간을 절약할 수 있고, 도표나 그래프와 같은 시각화 기능을 통해 분석이 더욱 용이해질 수 있다.

기업의 측면에서는 현재 상영중인 영화가 실제로 대중들에게 어떤 평가를 받고 있는지에 대한 신뢰도 높은 정보를 얻을 수 있으며, 이로 인해 오프라인이나 온라인으로 진행되는 마케팅 리서치에 들어가는 시간과 비용을 절약할 수 있다. 영화사의 경우 다른 경쟁사의 작품과 비교했을 때 어떤 속성에서 자사의 작품이 우위나 열위를 갖고 있는지를 쉽게 파악할 수 있고, 일반 마케팅 기업의 경우 대중에게 인기가 많은 영화에 대한 정보를 빠르게 습득하여 관련 마케팅 상품을 제작하는데 반영할 수 있다.

### 6-3. 사회·문화적 측면

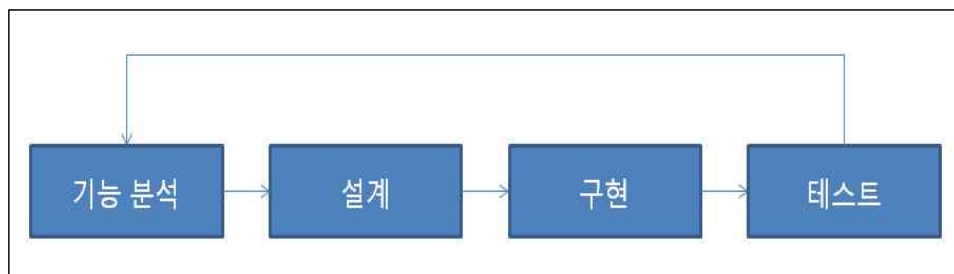
청소년부터 성인에 이르기까지 영화를 좋아하는 마니아는 많지만 이들을 위한 활성화된 커뮤니티 사이트는 아직까지 마땅한 곳이 없다. 주로 네이버 영화(<http://movie.naver.com>)의 한 줄 리뷰를 많이 사용하는데 다수의 사람들이 서로 대화를 주고 받기에는 적합하지 않은 공간이다. 만약 본 프로젝트 결과물이 대중으로부터 좋은 평가를 받게 되고 널리 알려진다면 서로가 영화에 대한 정보를 공유할 수 있는 새로운 커뮤니티 사이트의 출현을 기대할 수 있다.

## 7. 프로젝트 수행 체계

### 7-1. 프로젝트 역할 분담 및 수행 방법

수행기간	담당자	역 할
2012. 9. 1 ~ 2012. 12. 21	이봉기	역할 분담, 일정 관리, DB 설계 및 관리, 기획
	이 호	SVN 서버관리, 연락담당, 평판 분석 모듈 개발
	차소윤	회의록 작성, 송·수신 모듈 개발 및 관리
	위석철	보고서 종합 및 편집, 평판 분석 모듈 개발

#### 7-1-1. 개발 방법론 및 개발 과정



[그림 37] Spiral Model

[그림 37]과 같은 흐름으로 Spiral Model에 의해 먼저 개발을 한 후, 개선할 점과 추가할 점을 분석해 다시 개발하는 과정을 반복한다.

#### 7-1-2. 회의 방법

- 주 1회 이상 주말을 이용하여 과반수 참석 하에 회의를 진행한다. 한 주 동안의 성과를 공유하며, 다음 한 주 간의 목표와 계획을 수립한다.
- 월 1회, 팀원들과 멘토와의 오프라인 미팅을 진행한다. 프로젝트를 진행하면서 궁금했던 사항에 대해 멘토에게 질문하고, 멘토로부터의 조언을 듣는다.
- 월 2회, 팀원들과 교수님과의 오프라인 미팅을 진행한다. 현재까지의 진행 상황, 문제점에 대해 보고하고, 앞으로의 진행 방향에 대한 조언을 듣는다.

## 7-2. 프로젝트 수행 일정

### ○ 2012-1학기

구 분	내 용	담당자	4월				5월				6월	
			1	2	3	4	1	2	3	4	1	2
1	검색어 입력 모듈	이봉기										
2	SNS 데이터 수집 모듈	이봉기										
3	데이터베이스 설계 및 구축	이봉기										
4	소켓 통신 구현	차소윤										
5	DB 연동	차소윤										
6	데이터 전처리 모듈	이 호 위석철										
7	긍·부정 사전 구축	이봉기										
8	감성문장 분류 모듈	이 호 위석철										
9	극성 분류 모듈	이 호 위석철										
10	강도 분석 모듈	이 호 위석철										
11	데이터 출력 모듈	이봉기										
12	모듈 최적화	전 원										
12	모듈 통합	전 원										
13	테스트 & 디버깅	전 원										

### ○ 2012-2학기

구 분	내 용	담당자	7월				8월				9월				10월				11월				12월	
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
1	감성문장 분류 모듈 개선	위석철 이 호																						
2	극성 분류 모듈 개선	위석철 차소윤																						
3	영화 형용사 분석	이봉기 위석철																						
4	UI 개선	이봉기 차소윤 이 호																						
5	광고 문구 제거 모듈	이봉기 위석철																						
6	영화 형용사 UI	차소윤 이 호																						
7	기계학습에 의한 감성문장 분류	이봉기 이 호																						
8	기계학습에 의한 극성 분류	위석철 차소윤																						
9	테스트 & 디버깅	전 원																						

## 7-3. 기술적 장애 요소 및 해결방안

### ○ 클라이언트 개발 효율성

본 프로젝트는 사용자가 웹 페이지를 통해 영화의 분석 결과를 조회하고 볼 수 있도록 하는 것을 목적으로 한다. 따라서 웹 페이지의 UX를 쉽게 개발할 수 있고 서버 연동을 편리하게 할 수 있는 방법이 필요해 Flex 프레임워크를 기반으로 개발하기로 했다.

### ○ 클라이언트와 서버 간 연동

클라이언트와 서버 간에 데이터를 주고받기 위해 처음에 시도했던 방법은 소켓 통신을 직접 사용해 개발이 쉽지 않았고, 서버에서 처리를 해야 할 일을 클라이언트가 부담하게 되는 단점이 있었지만, 서버를 서블릿을 사용해 개발하고 HTTP로 통신하는 방법을 사용해 좀 더 쉽게 개발할 수 있었고 안정적인 연결을 구현했다.

### ○ 클라이언트와 DBMS 간 연결

분석 결과 조회를 구현할 때 원래 시도했던 방법은 클라이언트가 서버에 데이터를 요청할 때마다 매번 서버가 DBMS와의 연결을 생성한 다음 데이터를 가져오고 연결을 끊어 조회 속도가 매우 느렸다. 그래서 서버가 실행되고 최초로 요청이 들어왔을 때 DBMS와의 연결을 생성한 다음 그 후부터는 그 연결을 계속 사용하는 방법으로 조회 속도를 높일 수 있었다.

### ○ 쿼리문의 문자셋

본 프로젝트에서 다루는 영화의 제목은 한글 제목인 경우도 많기 때문에 쿼리문에 한글이 포함되어야 한다. 따라서 DBMS에 쿼리문을 보낼 때 쿼리문의 문자셋을 UTF-8로 해야 한다.

### ○ 문장의 상황에 따라 달라지는 감성단어 인식

평범한 단어가 부정적이거나 긍정적인 의미를 가질 수도 있다. 예를 들어 국회의원이 "오렌지"라는 말을 잘못 사용하여 국회의원이 오렌지라는 용어로 비판을 듣는다고 가정한다. 그럼 "오렌지"라는 의미는 감성단어가 아니지만 국회의원과 같이 쓰이면 부정적인 단어가 될 수 있다. 이런 내재적인 의미로 쓰이는 단어들을 구별하는 것은 어려운 난제이지만 해결해야 한다. 현재 본 프로젝트는 이 방식을 해결하지 못했고 이와 같은 방식을 해결하려면 구문분석기와 같은 프로그램이 필요할 것이라고 예상된다.

### ○ 문장의 상황에 따라 달라지는 감성단어의 극성

특정 감성단어는 문장의 상황에 따라 긍정의 의미를 가질 수도 있고 부정의 의미를 가질 수도 있다. 예를 들어 "무섭다"라는 용어를 공포영화에서 사용하면 긍정적인 의미의 단어가 된다. 하지만 살인마에게 "무섭다"라는 용어를 쓰면 부정적인 의미를 가지게 된다. 이와 같은 상황에 따라 감성단어의 극성은 바뀔 수 있다. 현재 본 프로젝트에서는 이 문제를 해결하지 못하였고 이 문제를 해결하기 위해서는 문장이 수집될 때, 검색 옵션의 상황을 미리 저장해놓는다면 약간의 해결책이 될 것이라고 예상된다. 또는 감성단어의 인식과 비슷한 문제이므로 구문분석기와 같은 프로그램이 필요할 것으로 예상된다.

## ○ 검색어와 관련된 문장의 미수집

검색어를 입력하여 데이터를 수집할 때, 검색어와 관련 있는 문장을 놓칠 수 있다. 띄어쓰기를 관점으로만 봐도 띄어쓰기가 되어져있는 검색어와 띄어쓰기가 되어져있지 않은 검색어는 같은 데이터이다. 하지만 수집하는 입장에서는 다른 문장으로 인식 될 수 있다. 따라서 이와 같은 문제를 해결하려면 모든 상황을 입력해주거나 그 상황을 예측해주는 인공지능을 가진 프로그램이 필요하다. 본 프로젝트에서는 이와 같은 문제를 아직 해결하지 못하였다.

## ○ 검색어와 관련되지 않은 문장의 수집

검색어를 입력하여 데이터를 수집할 때, 검색어와 관련 없는 문장이 나올 수 있다. 예를 들어 "돈의맛"이라는 영화를 검색하면 돈이나 맛과 관련된 데이터가 추가로 나올 수 있다. 이와 같은 경우 데이터를 가공하여 영화와 관련되지 않은 내용은 제거해야한다. 하지만 이와 같은 문제 또한 해결하지 못하였다. 영화와 관련 없는 특정 키워드를 가진 데이터는 제외시키는 기능을 추가할 경우 완전한 해결은 아니더라도 많은 개선이 보일 것이라고 예상된다.

## ○ 대화체 문장 처리

SNS로부터 얻는 개인의 의견들은 주로 비형식적인 대화체로서 띄어쓰기가 제대로 되어있지 않고 외래어, 줄임말, 이모티콘 등 기존의 형태소 분석기로는 분석하기 어려운 성분들로 이루어져있다. 띄어쓰기는 국민대 형태소 분석기에서 제공하는 모듈을 활용하여 해결할 수 있지만 외래어나 줄임말, 이모티콘의 경우는 현재까지는 이용 가능한 기술이 없으므로 개발자가 직접 데이터베이스에 관련 어휘를 삽입하여 신뢰도를 높이는 방법을 하면 개선될 것이라고 예상된다. 하지만 현재 본 프로젝트에서는 띄어쓰기문제만 해결하고 외래어나 줄임말, 이모티콘의 경우는 해결하지 못하였다.

## ○ 사전 구축

사전은 신뢰도를 결정짓는 주요한 성분 중 하나이다. 그렇기 때문에 사전은 이미 존재하는 수많은 어휘들을 포함하고 있어야 하며 새로운 신조어 또한 스스로 학습하는 기능이 필요하다. 하지만 이것들을 구현하기 위한 시간이 부족하기 때문에 감정단어를 제공하는 논문과 본 프로젝트를 개발하는 과정에서는 개발자가 직접 필요한 어휘를 추가로 설문조사하여 자료를 삽입하였다. 감성단어를 제공하는 논문에는 434개의 감성단어와 SNS데이터 속에서 기존 논문 감정단어사전에 없는 100개의 감성단어를 추가로 찾아내어 입력하였다. 100개의 감성단어는 54명의 사람들에게 설문조사를 시행하였다. 설문조사의 결과를 바탕으로 논문 감성단어사전에 맵핑을 하는 방식으로 감정수치의 문제를 해결하였다.

## ○ 형태소 분석기

형태소 분석기는 본 프로젝트를 원활히 수행하려면 반드시 구현해야 하는 기술이다. 하지만 형태소 분석기 개발은 그 자체만으로도 하나의 장기 프로젝트가 될 수 있는 만큼 복잡하고 어려운 분야이다. 그래서 본 프로젝트에서는 형태소 분석기를 직접 구현하지 않고 이미 개발되어 있는 형태소 분석기(국민대·창원대)를 사용한다.

## ○ 제목의 띄어쓰기

띄어쓰기 모듈이 제목을 마음대로 띄어쓰기 해버리는 경우가 있다. 예를 들어 제목이 ‘다크나이트’인 경우 띄어쓰기 모듈이 제목을 인식하지 못해 제멋대로 ‘다크나 이트’로 바꿔버리는 경우이다. 이런 경우를 위해 띄어쓰기 모듈에서 특정한 단어는 띄어쓰기를 적용시키지 않는 설정을 할 수 있다.

## ○ 광고 문구 제거

개인의 의견이 아닌 이벤트 및 광고성 문구들은 분석 결과의 신뢰도를 하향시키는 주된 요인이다. 하지만 그런 문구들은 SNS상에서 다양한 형태로 나타나기 때문에 개인의 의견으로부터 분리해내기가 어렵다.

SNS상에 같은 내용의 문구가 여럿 올라와 있는 경우 광고 문구로 간주하고 제거하는 모듈을 구현하였지만, 이 방법으로도 모든 광고 문구가 제거되는 것은 아니기 때문에 앞으로도 광고 문구를 제거하는 문제는 계속해서 연구할 필요성이 있다.

## ○ 제목에 포함된 형용사

제목은 명사로 간주되어 선호도 계산에 영향을 미치면 안 되지만, 제목에 포함된 형용사가 전체 문장의 선호도를 계산할 때 영향을 미칠 수 있다. 따라서 문장에서 제목을 제외하고 선호도를 계산하는 방법을 마련해야 한다.

## ○ 은어나 줄임말을 사용한 문장

SNS에서 흔히 등장하는 단어들은 주로 줄임말이나 인터넷 은어이다. 이런 단어들은 매우 다양한 형태로 등장하기 때문에 본 시스템에서는 이들 모두를 인식하여 분석을 수행하기가 상당히 어려운 상황이다. 그래서 본 프로젝트에서는 유사도 기법을 활용하여 비슷한 단어(예. '재밋다'와 '재있다')를 추출하고 이를 매핑시키는 방법을 통해 이와 같은 문제로 인한 성능 저하를 방지하였다.

#### 7-4. 기자재, 시설 및 장비 활용 현황

기자재 / 시설 / 장비명	규격	수량	용도	기보유 / 구매
노트북		4대	프로젝트 구현	4 / 0
데스크탑 / 정보과학관 304호		1대		1 / 0

#### 7-5. 참여 인력 현황

이름	프로젝트 경험	소속
이봉기	2011년 4월~6월, 안드로이드앱 "Smart Receiver" 개발 2011년 9월~12월, (주)자유에듀테인먼트 웹개발 인턴십	송실대학교 컴퓨터학부
이 호	2011년 5월~8월, "당진군 경찰서 통합관제시스템" 개발 2011년 1월~4월, 당진군청사 강당 전자플랜카드 "IMAB" 개발	
차소윤	2011년 IT대학 소프트웨어 공모전 "Cheese and Mice" 출품	
위석철	2012년 1월~4월 삼성 bada 공모전 "Smart Mouse" 개발	



## 8. 수상 현황

수상명	수상 내용	발급기관	수상 일자
2012 캡스톤 디자인 경진대회	금상	송실대학교	2012. 09. 20
2012년도 IT대학 소프트웨어 공모전	은상	송실대학교 IT대학	2012. 10. 10
한이음 엑스포 2012 작품 공모전	우수상	삼성SDS주식회사	2012. 10. 31

## 9. 참고 문헌

- [1] 강한훈, 유성준, 한동일, "상품 리뷰의 속성별 긍정/부정 분류 시스템의 설계 및 구현", 2009, 한국정보과학회 가을학술발표논문집 Vol.36, No.2(C)
- [2] 박건우, 임선열, 이태양, 김정훈, 최웅철, "감성 단어 분석 모듈을 이용한 모바일 소셜 네트워크 서비스에서의 성향 판단 시스템 구현", 2011, 한국정보과학회 가을학술발표논문집 Vol.38, No.2(D)
- [3] 양정연, 명재석, 이상구, "상품 리뷰 요약에서의 문맥 정보를 이용한 의견 분류 방법", 2009.8, 한국정보과학회 정보과학회논문지 제36권 제4호(pp.254-262)
- [4] 조상현, 강행봉, "형식적 및 비형식적 어휘 정보를 반영한 문장 감정 분류", 2011.10, 한국정보처리학회 정보처리학회지 제18-B권 제5호(pp.325-332)
- [5] 김현중, 조성준, "관객의 관람평 분석을 통한 영화의 특성 추출", 2012.4, 2012 한국BI데이터마이닝학회 춘계 학술대회
- [6] 송종석, 이수원, "상품평 극성 분류를 위한 특징별 서술어 긍정/부정 사전 자동 구축", 2011.3, 한국정보과학회 정보과학회논문지 제38권 제3호(pp.157-168)
- [7] 박경미, 박호건, 김형곤, 고희동, "SNS에서 오피니언 마이닝 연구", 2011, 한국정보과학회 정보과학회논문지 제29권 제11호(pp.54-60)
- [8] 임수민, 김형중, 주상현, "SNS에서 통용되는 User Data의 정의 및 표준화 동향", 2011.8, Proceedings of the ITFF Summer Conference(pp.361-363)
- [9] 박인조, 민경환, "한국어 감정단어의 목록 작성과 차원 탐색", 2005, Korean Journal of Social and Personality Psychology Vol.19, No.1(pp109~129)

## 첨부 A. 회의록

프로젝트명	SNS 텍스트데이터를 이용한 전자제품 평판 분석
일 시	2012년 3월 31일 12시 30분
장 소	정보과학관 3층 테라스
책임자	이봉기
목 적	기능 분석, 설계, 수행 일정 결정
참석자	이봉기, 차소윤, 위석철, 이호
기록자	차소윤
내 용	<ul style="list-style-type: none"> <li>- 미리 소분류를 소분류사전에 저장한 후 검색된 제품의 유형에 따라 소분류를 선택해 그 소분류에 대한 호불호를 분석하는 기능을 추가한다.</li> <li>- 검색어를 입력 받고 결과를 출력하는 모듈과 분석하는 모듈을 따로 만든다.</li> <li>- 1차적으로 분석 기능을 위주로 구현하고 2차에서 사용하기 좋게 UI를 구현하는 방향으로 한다.</li> </ul>
결정 사항	다음 회의 일시 : 2012년 4월 8일 6시