한글 감성어 사전 API 구축 및 자연어 처리의 활용

Building a Korean Sentiment Dictionary and Applications of Natural Language Processing

저자
(Authors)

안정국, 김희웅

출처
(Source)

한국지능정보시스템학회 학술대회논문집 , 2014.11, 177-182(6 pages)

발행처
(Publisher)

한국지능정보시스템학회
Korea Intelligent Information Systems Society

URL

http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06240876

APA Style

안정국, 김희웅 (2014). 한글 감성어 사전 API 구축 및 자연어 처리의 활용. 한국지능정보시스템학회 학술대회논문집, 177-182

한글 감성어 사전 API 구축 및 자연어 처리의 활용

Building a Korean Sentiment Dictionary and Applications of Natural Language Processing

안정국[a], 김희웅[b]

[a,b] *연세대학교 정보대학원*

서울특별시 서대문구 신촌동 134

Tel: 02-2123-4195, E-mail: [a]jace@yonsei.ac.kr, [b]kimhw@yonsei.ac.kr

초록

최근 데이터 빅뱅 시대로의 돌입과 기하급수적인 디지털 데이터의 활용과 더불어 자연어 처리가 각광을 받고 있다. 이에 본 연구는 한글 자연어 처리의 개발과 연구에 활용이 가능한 한글 자연어 처리 API를 구축하였다(오픈한글, www.openhangul.com). 한국어는 다른 언어에 비해 어미나 조사와 같은 문법적 형태가 발달한 교착어이므로 자연어 처리가 어려워 정보화나 정보시스템에의 활용이 미흡한 실정이다. 본 연구는 총 517,178(+)의 국어 단어에서 감성을 표현할 수 있는 명사, 형용사, 동사, 부사 등과 같은 단어 형태들을 중심으로 집단지성을 이용한 폭소노미 기반의 감성어 사전을 구축하는 프로젝트를 진행하였다. 또한 감성어 사전의 실무에서의 활용을 위해 API를 구축하여 문서의 긍정/부정/중립 감성 분석 및 자연어 처리에 응용이 가능하다. 테스트 결과 집단지성이 판단한 단어의 의미와 사전적 의미에 미묘한 괴리와 같은 새로운 발견하였으며 이는 본 연구에서 구축한 감성어 사전이 시간적인 개념을 반영하기 때문이다. 또한 감성의 깊이를 반영하는 특징도 있어 본 연구가 국어학과 공학분야 뿐만 아니라 융합학문적으로도 큰 의의가 있다고 본다. 본 연구에서 집단지성으로 구축한 한글 감성어 사전과 이를 비롯 다양한 오픈API를 제공을 하는 새로운 시도가 향후 한글 자연어 처리의 발전에 새로운 방향과 시사점을 제시를 할 수 있을 것이라 기대한다.

Keywords:

감성어 사전, 한글자연어처리, 감성분석, 폭소노미, 오픈API, 집단지성, 데이터마이닝, 텍스트마이닝, 크라우드소싱, 빅데이터

Ⅰ. 서론

최근 모든 분야에서 다양한 데이터의 급증으로 데이터의 중요성과 활용에 관한 부분이 관심을 받고 있으며 빅데이터환경에서 데이터 마이닝, 텍스트 마이닝, 오피니언 마이닝과 같은 것을 이용한 분석 기법들도 각광을 받고 있다. 또한 이러한 분석들의 목적은 기존에 간과했던 데이터간의 패턴을 발견하고 새로운 인사이트를 찾는 것에 많은 초점을 두고 있다. 빅데이터의 4Vs(Volume, Velocity, Variety, Value)(McAfee, 2012)에서 정의하듯이 다양하고 많은 데이터를 가져오는 것이 데이터분석의 시작이라고 할 수 있는데 현재로서는 기업들의 기술적 역량과 투자에 따라 차이를 보이고 있다. 하지만 최근 오픈 데이터를 활용한 개발, 공유가치, 협업등과 같은 다양한 개념이 대중들에게 호응을 받게 되고 새로운 유행이 됨에 따라 기업들도 이미지 개선과 기업중심의 생태계로 이끌어 내는 효과를 보고 있다. 이러한 기업들의 전략과 더불어 많은 기업들과 기관들이 다양한 데이터들을 API방식으로 xml이나 json과 같은 구조화된 데이터로 제공이 되기 때문에(Tu, 2009) 데이터의 수집에 대한 경쟁우위의 차이는 점점 작아지고 있다. 이러한 시점에서 기업들의 경쟁력은 데이터 분석에서 많은 차이를 보이게 될 것이며 빅데이터 환경에서의 데이터 분석이란 기존보다는 진보된 자연어 처리 등을 이용한 고급적인 분석을 의미한다.

최근 자연어 처리가 다양한 분야에서 각광을 받고 있는데 자연어 처리란 컴퓨터가 언어를 이해하여 분석하는 것을 말한다. 자연어 처리 중에서도 감성분석을 다양한 부분에서 활용하고 있으며 이는 인지공학적인 측면에서도 굉장히 중요한 역할을 하고 있다. 이와 같이 실무에서는 경영, 사회과학, 자연과학 등의 다양한 영역에서

데이터 마이닝과 텍스트 마이닝을 활용하여 사회현상, 트렌드, 브랜드, 제품, 마케팅, 여론 등과 같은 분석을 하는 추세이며 더 나아가 기업들이 미래에 대한 전략을 세우고 대비를 하는데도 활용을 한다. 하지만 정확한 분석을 위해서는 고차원적인 자연어 처리 기술과 더불어 신뢰도가 높은 감성어 사전이 필요한데 현재는 한국어 감성어 사전은 소수의 데이터 분석기업에서 구축을 한 상태다. 가장 큰 단점은 비용이 많이 들어 다수가 아닌 개인이 한 단어의 감성을 판단하는 방식이므로 개인적인 선입견(bias)이 들어가며 회사의 경쟁적 자원이므로 개방이나 공유를 하지 않는 것이다. 물론 회사의 입장에서는 당연하지만 한글 자연어 처리 기술의 발전에 있어서는 이러한 상황적인 부분이 큰 걸림돌이 된다는 의미이다.

또한 본 연구에서는 집단지성 외에도 사람들에 의한 분류를 하는 폭소노미(Folksonomy)를 적용하였다. 폭소노미는 Folks와 Taxonomy의 합성어로 시스템 설계사인 Thomas Vander Wal이 2004년에 만들었으며 협업적 태깅(Collaborative tagging), 소셜분류(Social classification), 소셜색인(Social indexing), 소셜태깅(Social tagging)의 의미로 실무에서는 쓰이고 있다(Wikipedia, 2014). 웹2.0이라는 개념과 더불어 블로그와 소셜커뮤니티 사이트들이 확산되면서 문서와 이미지에 태깅을 달아 디렉토리가 아닌 태그로도 카테고리 구분이 가능하게 하는 개념이다. Ohmukai, et al. (2006)는 폭소노미 기반으로 메타데이터(Metadata)를 이용한 소셜 북마크 온톨로지 구축을 제안하기도 하였으며 Medelyan & Legg (2008)은 온톨로지를 폭소노미에 활용하는 연구를 하기도 하였다.

폭소노미는 다양한 정보를 얻을 수 있지만 태깅이 무작위이므로 신뢰성이 떨어진다는 단점이 있다(Gruber, 2005). 실제로 이러한 한계점으로 인해 폭소노미에 대한 관심과 연구는 2007년 이후로 많이 감소된 추세이다. 기존 연구들을 보면 새로운 태깅과 클러스터링을 하는 관점이며 실질적인 구축이 아닌 제안하는 수준이다.

본 연구의 연구자들은 폭소노미의 특성인 참여에 의한 협업과 분류가 빅데이터 환경의 시스템에 근본적인 틀과 방향성을 줄 수 있을 것이라고 확신을 하였고 폭소노미의 단점인 신뢰성을 극복하기 위해 집단지성(Collective intelligence)(Malone, Laubacher, and Dellarocas, 2010)을 활용한다. 이는 개인적인 선입견이 들어가지 않는 감성어 사전의 구축이 궁극적 목적이기 때문이다. 이와 더불어 API (Application programming interface)를 개방하여 누구나 연구나 실무에서의 활용이 가능하게 하고자 하여 다양한 분야 연구자들의 학문적인 참여와 업계의 실무적인 참여를 이끌어 한글 자연어 처리의 협력적 발전을 도모하는 데 목적을 가지고 있다.

Ⅱ. 연구설계

본 연구는 (그림 1)처럼 집단지성으로 감성어 사전을 구축을 하며 (그림 2)와 같은 텍스트 마이닝 감성분석 프로세스를 기본적인 틀로 잡았다. 감성어 사전을 구축하기 위한 전처리 작업으로 국어사전 데이터베이스를 구축하였고 국립국어원의 사전을 기반으로 총 517,178(+)개 단어를 데이터베이스에 생성하였다. 마지막 단계에서는 감성분석에서 감성어 사전이 쓰이며 이를 다시 오픈API로 개인, 연구자, 기업들에게 제공을 한다 (그림 1).

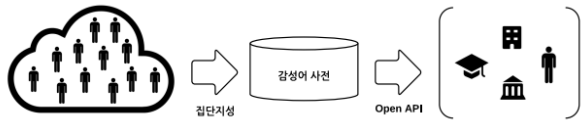


그림 1 - 집단지성의 결과물을 오픈API로 전달

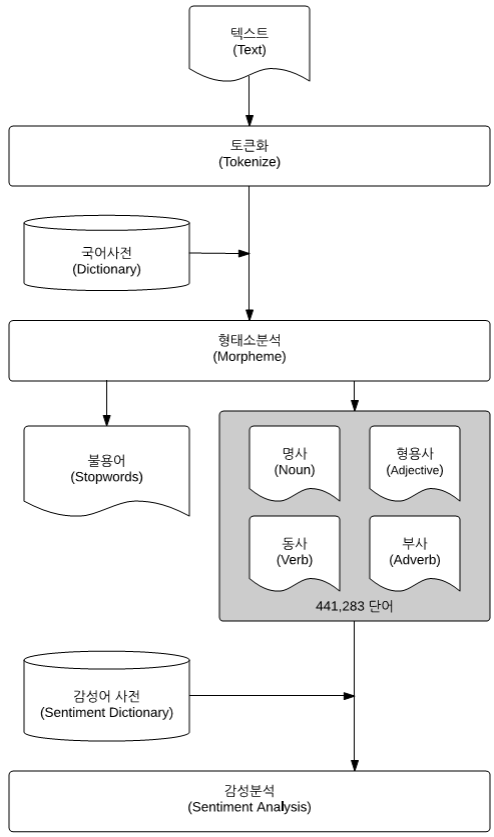


그림 2 – 텍스트 마이닝 프로세스

2-1 사전데이터 설계

기본 색인 구축에 있어 1차 카테고리는 단어의 'ㄱㄴㄷ'의 모음 순으로 하였고, '가나다'의

자모 결합 순으로 2차 카테고리로 나누었다 <표 1>. 추가로 자모를 분리하여 저장하였는데 이는 단순히 분류나 나열을 위한 것이 아니라 검색의 최적화, 유사도, 오타인식, 기본형 추출을 위한 것이다. 본 연구에서는 편집거리(Edit distance)를 기반으로 하는 레빈쉬타인 거리(Levenshtein Distance) 매트릭스<그림 3>, (1)를 검색에 사용하며 삭제/삽입/수정 (deletions/insertions/reversals)의 횟수의 총합으로 계산을 하므로 순차적으로 인식하는 방식과 다르다 (Levenshtein, 1966). 예를 들면 'Today'와 'Tonight'의 편집거리는 5이다. 알파벳의 경우 문자당 1byte인 반면 한국어는 초성, 중성, 종성의 결합형인 2byte이므로 자음과 모음을 각각 1byte로 인식하도록 분리하여 데이터베이스에 입력을 하였다 <표 1>. 즉 커서의 이동단위를 자모 단위로 가능하도록 하여 레빈쉬타인 거리 계산을 용이하게 한다.

$$\mathrm{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \mathrm{lev}_{a,b}(i-1,j)+1 \\ \mathrm{lev}_{a,b}(i,j-1)+1 \\ \mathrm{lev}_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \quad (1)$$

		k	i	t	t	e	n
	0	1	2	3	4	5	6
s	1	1	2	3	4	5	6
i	2	2	1	2	3	4	5
t	3	3	2	1	2	3	4
t	4	4	3	2	1	2	3
i	5	5	4	3	2	2	3
n	6	6	5	4	3	3	2
g	7	7	6	5	4	4	3

		S	a	t	u	r	d	a	y
	0	1	2	3	4	5	6	7	8
S	1	0	1	2	3	4	5	6	7
u	2	1	1	2	2	3	4	5	6
n	3	2	2	2	3	3	4	5	6
d	4	3	3	3	3	4	3	4	5
a	5	4	3	4	4	4	4	3	4
y	6	5	4	4	5	5	5	4	3

<그림 3> 레빈쉬타인 거리 매트릭스의 예

<표 1> 사전의 필드와 값의 예

ㄱ	거	거머릿과	ㄱㅓㅁㅓㄹㅣㅅㄱㅘ	명사
ㄱ	거	거머먹다	ㄱㅓㅁㅓㅁㅓㄱㄷㅏ	동사
ㄱ	거	거머멀쑥	ㄱㅓㅁㅓㅁㅓㄹㅆㅜㄱ	
ㄱ	거	거머멀쑥이	ㄱㅓㅁㅓㅁㅓㄹㅆㅜㄱㅇㅣ	부사
ㄱ	거	거머멀쑥하다	ㄱㅓㅁㅓㅁㅓㄹㅆㅜㄱㅎㅏㄷㅏ	형용사
ㄱ	거	거머무트룩하다	ㄱㅓㅁㅓㅁㅜㅌㅡㄹㅜㄱㅎㅏㄷㅏ	형용사
ㄱ	거	거머무트름	ㄱㅓㅁㅓㅁㅜㅌㅡㄹㅡㅁ	
ㄱ	거	거머무트름하다	ㄱㅓㅁㅓㅁㅜㅌㅡㄹㅡㅁㅎㅏㄷㅏ	형용사
ㄱ	거	거머무트름히	ㄱㅓㅁㅓㅁㅜㅌㅡㄹㅡㅁㅎㅣ	부사
ㄱ	거	거머번드르	ㄱㅓㅁㅓㅂㅓㄴㄷㅡㄹㅡ	
ㄱ	거	거머번드르하다	ㄱㅓㅁㅓㅂㅓㄴㄷㅡㄹㅡㅎㅏㄷㅏ	형용사

2-2 우선 순위 단어의 형태 선별

총 517,178(+)개 단어 중에서 감성의 표현이 가능한 단어의 형태들을 선별하였으며 조사, 감탄사, 관형사 등의 단어들은 제외를 하고 명사, 형용사, 동사들을 우선순위로 두었다 <표 2>. 그리고 동의어의 경우에는 첫 번째 순위에 있는 의미를 선택하는 단어의 중요도를 기반으로 우선순위를 두었다. 또한 불용어로 분류된 단어 형태에도 감성 단어들이 간혹 있기는 하지만 전체적인 변별력을 높이기 위해 감성어 대상에서 제외하였는데 이는 감탄사의 경우만 하더라도 문장의 성격과는 무관하게 주변 단어의 강조를 위해 쓰이는 경우가 많기 때문이다. 총 517,178개의 단어 중 불용어로 처리된 단어들은 75,895개이다 <표 2>.

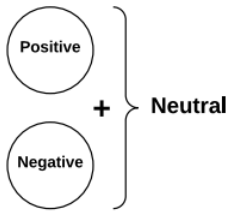
<표 2> 단어형태 분류

단어형태	단어 수	분류
감탄사	682	불용어
감탄사 · 명사	85	불용어
관형사	207	불용어
관형사 · 감탄사	11	불용어
관형사 · 명사	1,267	감성어*
대명사	382	불용어
대명사 · 감탄사	5	불용어
대명사 · 관형사	3	불용어
대명사 · 부사	1	불용어
동사	68,370	감성어*
동사 · 형용사	2	감성어*
명사	337,659	감성어*
명사 · 부사	109	불용어
보조동사	14	불용어
보조형용사	17	불용어
부사	17,425	감성어*
부사 · 감탄사	3	불용어
수사	60	불용어
수사 · 관형사	195	불용어
수사 · 관형사 · 명사	3	불용어
어미	6	불용어
의존명사	913	불용어
의존명사 · 조사	13	불용어
접사	209	불용어
조사	300	불용어
형용사	16,562	감성어*
총 단어 수	517,178	불용어+감성어
총 감성어	441,283	감성어

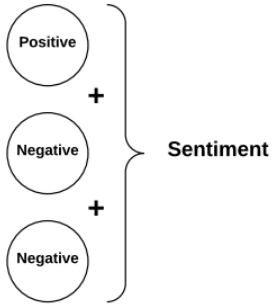
2-3 단어들의 감성 태그 및 카테고리 분류

태그는 성격에 따라 유형의 구분이 된다. Xu, et al. (2006)은 태그를 Content-based, Context-based, Attribute, Subjective, Organizational로 5가지로 분류를 하였는데 본 연구의 감성 태그는 단어의 내용에 대해 표현하는 '내용 기반형'이며 긍정, 부정, 중립의 판단과 감성의 깊이가 태그가 된다. 이는 폭소노미의 자율적인 태깅과는 다른 객관식 방식의 선택적 태깅인데 기존의 폭소노미의 단점인 태깅의 남용(Gruber, 2005)으로 인한 낮은 신뢰도와 정확도를 극복하기 위한 것이다. 단어 별로 사람들이 긍정, 부정, 중립을 선택을 하게 하여 이를 기반으로 감성의 깊이나 확률을 정량적으로 계산한다. 즉 세

가지의 제한된 태깅을 하지만 표현은 개인의 판단이 아닌 집단지성의 합이므로 단어마다 감성의 깊이 표현이 가능하다. 예를 들자면 A라는 사람이 '시원하다' 의 단어에 긍정의 태깅을 선택한다면 100% 긍정어로 태깅이 되는 것이 아니라 기존의 집단지성으로 계산된 값에 추가가 되어 A의 긍정 태깅이 부분적으로 반영이 되는 것이다. 이런 방식의 장점은 태깅의 남용을 방지하고 신뢰성을 높일 수 있다. 또한 언어는 시간에 따라 조금씩 변하는데 본 연구는 과거와 현재를 반영하므로 언어의 시간적 변화를 포함한다. 감성 점수를 계산하는 방법은 중립을 어떻게 처리를 하냐에 따라 두 가지 방법을 사용한다. <그림 4>처럼 긍정과 부정의 점수만 계산을 하여 중립을 자동으로 분류를 하는 방법과 긍정, 부정, 중립을 독립적으로 계산하는 방법<그림 5>이 있다. 전자의 경우에는 중립으로 수집된 태그들을 무시를 하는데 그 이유는 사람들이 단어의 뜻을 모를 경우에도 중립으로 태깅을 하기 때문이다. 그래서 감성 단어형태에서도 불용어를 2차적으로 분리할 수 있는 장점이 있다 <표 2>. 점수화 계산은 단어의 긍정 또는 부정일 확률이 0% ~ 60% (threshold=60%)이면 자동적으로 중립으로 분류가 되며 긍정이나 부정으로 판별이 되기 위해서는 최소한의 60% 이상의 확률이 있어야 한다. 후자의 긍정, 부정, 중립을 독립적으로 점수화한 방식은 비감성어도 중립어로 인식을 하게 하였으며 실질적으로 비감성어와 중립어의 차이는 있지만 텍스트 마이닝 분석에 있어서는 별 차이는 없다. 본 연구에서 쓰인 감성어 점수 알고리즘은 첫 단계에는 단어의 중립을 판단하기 위해 전자, 두 번째 단계에서 긍정과 부정을 확률을 계산을 위해 후자의 방법을 쓴다 <그림 5>.



<그림 4> 긍정, 부정으로 중립결정



<그림 5> 긍정/부정/중립으로 감성을 정량화

```
//중립 확률 계산 공식
if($data[neutral] > $data[positive] && $data[neutral] > $data[negative])
{
        ${score.$i} = 100 * $data[neutral]/(abs($data[positive] -
$data[negative]) + $data[neutral]);
        ${sentiment.$i} = '중립';
}
//긍정 확률 계산 공식
elseif($data[positive] > $data[negative]) {
        ${score.$i} = 100 * $data[positive]/($data[positive] +
$data[negative]);
        ${sentiment.$i} = '긍정';
}
//부정 확률 계산 공식
elseif($data[positive] < $data[negative]) {
        ${score.$i} = 100 * $data[negative]/($data[positive] +
$data[negative]);
        ${sentiment.$i} = '부정';
}
// 긍정/부정/중립의 수가 같을 경우
else {
        ${score.$i} = '100';
        ${sentiment.$i} = '중립';
}
```

<그림 6> 감성 점수계산 알고리즘

2-4 데이터 수집

데이터 수집을 위해 객관식 설문문항 형식으로 국내 대학생 소셜네트워크 사이트에서 실시하였으며 응답자들은 주어진 단어들에 대해 중립, 긍정, 부정의 보기 중에 택일을 하였다. 예를 들면 '가난하다'라는 단어가 주는 느낌이 중립, 긍정, 부정인지를 판단을 하는 것이다. 본 연구의 집단지성에 참여하는 '집단' 에 해당되는 구성원들의 정의는 기본적인 교육적 소양을 가진 다양하고 평범한 사람들이며 대학생들이 다른 집단보다 본 프로젝트에 적절하다고 생각했다. 응답오차 방지를 위해 의도적인 문항을 두 개로 두어 에러를 줄였다. 예를 들면 '좋다' 와 같은 단어를 부정어라고 선택한 경우 나머지 문항의 태그들도 오류로 인식을 하여 태깅에 반영을 하지 않는다. 웹 서비스의 특성과 편의성을 고려하고, 답변의 정확도를 높이기 위해 응답시간이 1분 내외로 걸리도록 하여 한 명당 총 10 단어씩 답변을 하게 하였다.

2-5 감성의 깊이 표현 알고리즘

본 연구의 감성 점수화 공식은 <그림 7>와 같으며 1단계에서 중립의 수가 긍정과 부정의 합이 아닌 각각의 수보다 많으면 중립의 단어로 판별한다. 그 후 2단계에서는 긍정 단어와 부정 단어의 수를 비교해서 %로 표현을 한다.

Ⅲ. 연구결과 및 실무 활용

본 연구에서 구축된 감성어 사전은 빅데이터 분석을

하는 회사나 기관들이 API를 이용하여 사용을 할 수 있도록 구현을 하였으며 감성 사전에서의 단어 질의 외에도 기본형 추출, 카테고리 추출과 같은 추가적인 API도 제공을 한다 (www.openhangul.com).

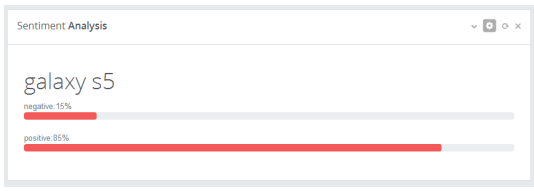
3-1 감성어 사전 (API)

집단지성으로 만들어진 단어의 감성 점수를 API로 받을 수 있다. 감성이라는 것이 주관적인 판단이므로 결과값에 대한 신뢰도의 판단은 연구자들의 주관적으로 해야 했으며 이를 보완하기 위해 단어들의 사전적 정의와 비교를 하였다. 이 부분에서 새롭게 발견한 부분은 연구자들의 주관적 판단과의 비교보다 사전적 정의와 비교를 하는 것이 좀 더 집단지성과 가까웠다는 것이다. 이는 감성 사전 구축에 있어서는 집단이 개인보다 더 적합하다는 것이다. <그림 8>처럼 '좋다' 라는 질의어 (query)를 요청하면 RESTful API 방식으로 실시간 웹으로 응답을 받을 수 있다.

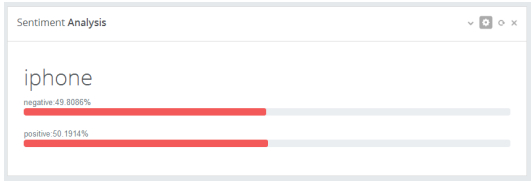

```
{
"word": "좋다",
"type": "형용사",
"sentiment": "긍정",
"sentiment_score": "99%"
}
```


<그림 8> 감성 단어 API의 출력결과의 예

3-2 텍스트 마이닝



<그림 9> 갤럭시 분석 (뉴스/SNS/댓글 분석)

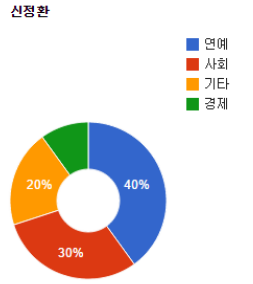


<그림 10> 아이폰 분석 (뉴스/SNS/댓글 분석)

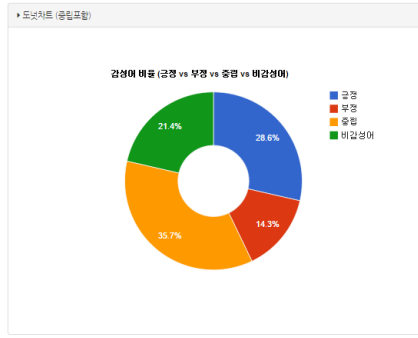
감성어 사전 API를 확장 활용하면 브랜드, 경쟁사, 평판 분석 등에 적용 가능하다. 또한 시간적 시점에 따른 다각적인 분석도 가능한데 회사가 신제품에 대하여 출시 전, 후에 대한 자사 제품의 평판을 분석을 할 수도 있고 반대로 자사 제품의 출시가 경쟁사의 평판에도 영향을 미치는 정도와 같은 부분이다. 부가적인 설명을 하자면 출시 전에 대한 분석은 기대치를 반영하는 preview이며, 출시 후는 고객의 반응을 반영하는 review가 될 수 있다. 예로 <그림 9>에서 갤럭시 S5에 대한 사람들의 반응은 85%의 긍정과 달리 아이폰은 50%만 긍정으로 나온다. 감성분석을 한 시점은 아이폰 6가 나오기 전의 시점이므로 사람들의 기대치가 낮다는 분석이 가능하며 확장적인 해석이 가능하다. 또한 '기대치'와 '사용후기'를 비교하여 '별로지만 기대했던 것보다는 덜 별로다'와 같은 주관적 분석도 가능하다.

3-3 키워드 카테고리 분석 (API)

<그림 11>처럼 키워드가 속한 현재의 카테고리를 분석을 하여 웹에서의 텍스트 정보를 기반으로 시간적인 개념을 반영을 하므로 시간에 따른 변화를 감지를 할 수 있는 특징이 있다.



<그림 11> 카테고리 분포 비율



<그림 12> 긍정/부정/중립/ 분포 분석

<그림 13> 긍정/부정으로만 감성 점수화

3-4

개발자들이나 연구자들의 텍스트 마이닝 분석 알고리즘은 주관적이므로 다양한 적용이 가능하다. 예를 들면 텍스트를 토큰화하여 <그림 12>처럼 중립을 포함을 하거나 <그림 13>처럼 중립을 제외를 하여 알고리즘에 적용을 할 수 있다.

Ⅳ. 연구결과에 대한 기대효과 및 활용방안

서론에서의 설명에 추가를 하자면 본 연구의 학문적, 실무적 적용과 활용영역은 <표 3>처럼 광범위하며 이 분야의 연구적인 측면도 다양한 분야의 융합을 요구한다. 예를 들면 어문학의 학문적 부분과 공학의 기술 구현, 그리고 경영에서의 실무와 같이 다양한 분야에서의 융합과 적용이 필요하다. 본 연구의 프로젝트를 진행하며 API를 제공한 결과 다양한 기업들과 연구자들이 적극적인 관심을 보이는 것을 확인하였다. 이는 한글 자연어 처리에 관한 연구나 개발의 필요성과 모티브는 많지만 실질적으로 한글 자연어처리의 구조적인 한계로 인해 연구나 개발에 제약이 있었던 것이다. 본 연구는 집단지성으로 구축된 감성어 사전과 같은 데이터를 개방을 하여 참여집단, 연구자, 개발자, 그리고 누구나 한글 자연어 처리 개발에 참여하여 주도하는 방식의 토대가 되기를 기대하는 바이다.

<표 3> 한글 자연어 처리의 활용영역

경영	정보화 처리, 브랜드/제품 모니터링, 마케팅 효과 측정, 뉴스/SNS/댓글분석, 브랜드/제품 모니터링, 마케팅 효과 측정, 여론측정, 기업 평판 리스크, 기업투명성 분석
자연과학	빅데이터, 자연어 처리, 데이터마이닝, 텍스트마이닝, 시맨틱웹, 온톨로지, 기계학습
사회과학/문과	사회현상, 트렌드 분석, 문헌정보, 텍스트마이닝, 언어학

References

Ducatel, G., & Virginas, B. (2009, July). A knowledge discovery tool for mobile worker support. In Computers & Industrial Engineering, 2009. CIE 2009. International Conference on (pp. 1390-1394). IEEE.

Echarte, F., Astrain, J. J., Córdoba, A., & Villadangos, J. E. (2007). Ontology of Folksonomy: A New Modelling Method. SAAKM, 289, 36.

Folksonomy. (n.d.). In Wikipedia. Retrieved October 12, 2014, from http://en.wikipedia.org/wiki/Folksonomy

Gruber, T. (2005). Folksonomy of ontology: A mash-up of apples and oranges.

Hwang, S. H., & Kang, Y. K. (2008). Hierarchical Triadic Context Analysis for Folksonomy-Based Web Applications. JDCTA, 2(1), 20-27.

Levenshtein, V. I. (1966, February). Binary codes capable of correcting deletions, insertions and reversals. In Soviet physics doklady (Vol. 10, p. 707).

Malone, T. W., Laubacher, R., & Dellarocas, C. (2010). The collective intelligence genome. IEEE Engineering Management Review, 38(3), 38.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big Data. The management revolution. Harvard Bus Rev, 90(10), 61-67.

Medelyan, O., & Legg, C. (2008, July). Integrating Cyc and Wikipedia: Folksonomy meets rigorously defined common-sense. In Proceedings of the WIKI-AI: Wikipedia and AI Workshop at the AAAI'08 Conference, Chicago, US.

Nicotra, J. (2009). "Folksonomy" and the Restructuring of Writing Space. Writing, 300, 61.

Ohmukai, I., Hamasaki, M., & Takeda, H. (2005, November). A proposal of community-based folksonomy with RDF metadata. In Proceedings of the 4th International Semantic Web Conference (ISWC2005).

Ohkura, T., Kiyota, Y., & Nakagawa, H. (2006, May). Browsing system for weblog articles based on automated folksonomy. In Proceedings of the WWW 2006 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, at WWW (Vol. 2006).

Russell, T. (2005). Contextual authority tagging: Cognitive authority through folksonomy. Unpublished manuscript. Retrieved, 11(16), 2005.

Trant, J., & Wyman, B. (2006, May). Investigating social tagging and folksonomy in art museums with steve. museum. In Proceedings of the WWW'06 Collaborative Web Tagging Workshop.

Tsai, Angela Charng-Rurng, et al. "Building a concept-level sentiment dictionary based on commonsense knowledge." *IEEE Intelligent Systems*, 28.2 (2013): 22~30.

Tummarello, G., & Morbidoni, C. (2007, May). Collaboratively Building Structured Knowledge with DBin: From del. icio. us Tags to an" RDFS Folksonomy". In CKC.

Tu, S. (2009). Exploiting linked data to build web applications.

Umbrath, A. S. R. W. W., & Hennig, L. (2009). A hybrid PLSA approach for warmer cold start in folksonomy recommendation. Recommender Systems & the Social Web, 10-13.

Vander Wal, T. (2007). Folksonomy. online posting, Feb, 7.

Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. Decision Support Systems, 57, 77-93.

Xu, Z., Fu, Y., Mao, J., & Su, D. (2006, May). Towards the semantic web: Collaborative tag suggestions. In Collaborative web tagging workshop at WWW2006, Edinburgh, Scotland.