오피니언 마이닝 기술을 이용한 효율적 상품평 검색 기법

(An Efficient Search Method of Product Reviews using Opinion Mining Techniques)

윤 홍 준 † 김 한 준 ††
(Hongjune Yune) (Han-joon Kim)

장 재 영 †††
(Jae-Young Chang)

요 약 급속한 전자상거래의 발전으로 인하여 온라인상으로 상품을 구매하고 그에 대한 평가를 작성하는 것이 일반적인 구매 패턴이 되었다. 구매자들의 상품평은 다른 잠재적인 소비자들의 상품 구입을 이끌어내는데 큰 동기가 된다. 하지만 온라인 쇼핑몰에서는 상품평의 성질에 부합하는 순위를 부여하지 않기 때문에, 사용자가 구입 결정을 위하여 수많은 상품평에 포함된 의견들을 효과적으로 검토하기는 쉽지 않다. 일반적으로 상품평은 감정적이며 주관적인 의견을 포함하고 있다. 그래서 이러한 상품평에 순위를 부여하는 방법은 일반 웹 검색과는 달라야 한다. 본 논문에서는 오피니언 마이닝 기술을 이용하여, 사용자의 의도에 따라 상품평 데이터에 대해 순위를 결정하는 기법을 제안한다. 제안된 기법은 사용자의 검색어뿐만 아니라 상품평 내에 주관적인 의견의 포함 여부 및 감정 극성의 엔트로피 등을 고려하여 상품평의 가치를 판단하였다. 또한 실험을 통하여 제안된 기법의 우수성을 검증하였다.

키워드 : 오피니언 마이닝, 오피니언 검색, 상품평, 감정 극성, 엔트로피

Abstract With the continuously increasing volume of e-commerce transactions, it is now popular to buy some products and to evaluate them on the World Wide Web. The product reviews are very useful to customers because they can make better decisions based on the indirect experiences obtainable through these reviews. However, since online shopping malls do not provide ranking results, it is not easy for users to read all the relevant review documents effectively. Product reviews include subjective and emotional opinions. Thus, the review search is different from the general web search in terms of ranking strategy. In this paper, we propose an effective method of ranking the reviews that can reflect user's intention by using opinion mining techniques. The proposed method analyzes product reviews with query words, and sentimental polarity of subjective opinions. Through diverse experiments, we show that our proposed method outperforms conventional ones.

Key words : Opinion Mining, Opinion Search, Product Review, Sentimental Polarity, Entropy

1. 서 론

전자상거래의 발달로 인하여 온라인상에서 소비자 간에 상품에 대한 정보를 공유하는 것은 일반적인 현상이 되었다. 웹 2.0시대인 오늘날에는 잠재적 소비자가 기존 구매자들의 상품에 대한 평가를 찾아내는 것은 어려운 일이 아니다. 여러 사용자들에 의해 상품평(product review)이 작성되고, 서로 공유되고 있기 때문이다. 이러한 상품평들은 잠재적인 소비자들의 구매 결정에 큰 역할을 하게 되지만, 사용자 수가 많은 만큼 상품평도 많이 존재한다. 그러므로 잠재적 소비자가 상품평을 능률적으로 검토하기 위해서는 적지 않은 노력이 필요하다. 이러한 문제에 도움이 되는 대표적인 기술로는 오피니언 마이닝(opinion mining)이 있으며, 현재 이 분야에 대한 다양한 연구가 진행되고 있다[1-3]. 오피니언 마이닝 기술은 사용자들의 상품평을 효과적으로 추출하여 요약함으로써 잠재적 소비자들로 하여금 모든 상품평을 살펴보지 않더라도 상품에 대한 다양한 의견들을 쉽게 열람 및 검색할 수 있는 기반을 제공한다.

본 논문에서는 오피니언 마이닝 기술을 이용하여 상품평을 검색하고자 하는 사용자의 의도에 따라 상품평들의 랭킹을 부여하는 기법을 제안한다. 제안된 기법의 특징은 검색어에 대한 TF(term frequency)뿐만 아니라 상품평의 감정 극성(sentiment polarity)에 대한 엔트로

· 본 연구는 2008년 서울형산업 기술개발 지원사업(과제번호: NT080624)의 지원을 받았으며, 또한 2009년도 한성대학교 교내연구비 지원을 받았음
· 이 논문은 2009 한국컴퓨터종합학술대회에서 '오피니언 마이닝 기술을 이용한 효율적 상품평 검색 기법'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 서울시립대학교 전자전기컴퓨터공학부
flow@uos.ac.kr
†† 종신회원 : 서울시립대학교 전자전기컴퓨터공학부 교수
khj@uos.ac.kr
(Corresponding author임)
††† 종신회원 : 한성대학교 컴퓨터공학과 교수
jychang@hansung.ac.kr

논문접수 : 2009년 8월 14일
심사완료 : 2009년 11월 10일

피(entropy)의 크기, 상품평 내 검색어와 감정 단어(sentiment word)의 근접도, 상품평 전체의 감정 단어의 TF를 바탕으로 상품평의 순위를 선정한다. 또한 본 논문에서 제안된 기법을 검증하기 위해서 실험을 실시하였다. 실험은 온라인 쇼핑몰인 Amazon.com의 상품평을 대상으로 하였으며, 웹 검색엔진의 검색 결과와의 비교 실험을 통해 검색 성능이 우수하다는 것을 증명하였다.

2. 관련 연구

텍스트 마이닝의 한 분야로서 최근에 오피니언 마이닝에 관한 연구가 활발히 이루어지고 있다. 특히 오피니언 마이닝의 가장 성공적인 응용분야가 온라인 쇼핑몰 환경에서의 상품평 분석이다. 그 이유는 일반적인 구매자의 상품평은 다른 분야의 텍스트들에 비해 주관적인 의견들이 보다 분명히 작성되어있기 때문이다.

상품평의 정확한 요약을 위한 여러 기법들에 대한 연구가 진행되고 있다[1-6]. 상품평을 요약하는 방법으로는 크게 자연어 처리기법과 통계학적 접근법이 있다[5]. 이 중 자연어 처리기법은 전문가의 시간과 노력이 많이 필요한 단점을 가지고 있어 이를 대신하여 통계학적 접근법이 주목받고 있다. [1]에서는 오피니언 마이닝에 대한 접근을 위해 문장 구조와 문장 사이의 관계, 문장성분의 패턴 정보 등의 언어 규칙을 이용한 통계학적 방법을 제시하고 있으며, [2]와 [4]에서는 WordNet을 활용하여 어휘의 긍정이나 부정적 의미를 판단하고, 이를 Senti-WordNet으로 응용하여 감정의 폭을 정량화하는 방법을 제시하고 있다. 이외에도 다양한 기법을 이용하여 상품평들을 분석하고자하는 노력들이 활발히 이루어지고 있다.

3. 상품평 검색 기법

3.1 상품평 검색 모델

상품평은 일반적인 문서보다 길이가 짧고 상품의 특성(feature)과 의견(opinion)이 많이 포함되어 있다. 예를 들어 디지털 카메라는 크기, 가격, 무게, 디자인, 배터리, 렌즈, 액정, 셔터스피드, 플래시 등의 특성을 가지고 있다. 사용자는 이러한 상품의 특성에 대한 감정적인 의견을 가질 수 있다. 카메라의 디자인에 대해서 <좋다/나쁘다>라는 기본적인 의견부터 액정이 <밝다/어둡다/선명하다/흐리다/크다/작다> 등의 다양한 의견까지 표현할 수 있다.

이러한 환경에서 사용자는 상품평 검색시 상품의 특성으로 검색하는 경향이 있다. 이는 사용자의 목적이 상품의 특성에 대한 평가를 알고자 하는 것이기 때문이다. 예를 들어 디지털 카메라에 대한 상품평을 검색하고자 하는 사용자는 '배터리'에 대한 긍정 혹은 부정적인 상품평을 검색하고자 할 것이다. 본 논문에서는 검색어로 상품에 대한 특성이 주어진다고 가정하며, 이를 바탕으로 해당 특성에 대해 가장 연관되며, 특히 주관적인 판단이 잘 포함된 상품평들을 우선적으로 선정하는 기법을 개발한다.

3.2 시스템 구조

본 논문에서 개발한 시스템의 전반적인 구조는 그림 1과 같다. 우선 크롤러(crawler)는 여러 사용자들이 작성한 온라인 쇼핑몰의 상품평들을 수집한다. 수집된 상품평은 구문 분석 작업을 통해 불용어 및 특수문자가 제거되고, 나머지 의미 있는 단어들은 데이터베이스에 저장된다. 추출된 단어는 상품평별로 저장되며 단어 중 미리 구축되어 있는 상품의 특성, 긍정 및 부정의 의미를 반영하는 단어 목록과 일치하는 단어는 따로 저장한다. 이를 바탕으로 상품평의 긍정/부정의 극성을 판단할 수 있도록 한다. 또한 상품평 내에서 단어들의 위치 정보를 같이 저장하여 추후 부가적인 정보를 알아낼 수 있도록 한다.

사용자가 상품의 특성이라고 할 수 있는 단어로 질의를 할 경우, 저장된 데이터베이스를 기반으로 하여 검색 및 순위 선정 모듈이 상품평에 적절한 조건을 부여하여 순위를 선정한다. 마지막으로 선정된 결과를 사용자에게 돌려주어 사용자가 능률적으로 상품평을 탐색할 수 있도록 한다.

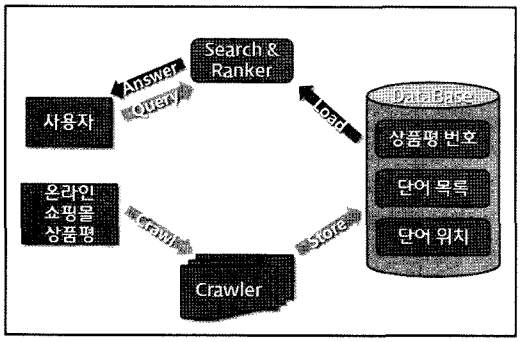

그림 1 시스템 구조도

3.3 상품평 검색 기법

본 논문에서는 상품평 검색 결과에 대한 우선순위 선정을 위해 일반적인 웹 데이터 순위 선정 방식과는 다른 몇 가지의 기준을 추가적으로 선정하였다. 우선 사용자가 질의한 단어의 TF가 큰 상품평 순서로 순위를 선정한다. TF는 문서 내에서 특정 단어가 출현하는 회수를 모든 단어가 출현하는 회수로 나눈 값으로 다음의 수식으로 표현된다.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

이 식에서 $n_{i,j}$는 문서 d_j에서 단어 t_i의 출현 회수를, $\sum_k n_{k,j}$는 문서 d_j에서 모든 단어의 출현 회수를 의미한다. 키워드의 TF로 순위를 선정하는 방법은 기존의 방식과 크게 다르지 않다. 하지만 상품평은 그 길이가 일반적인 웹 페이지 만큼 길지 않으므로, 검색어에 대한 TF가 동률일 상황이 많이 생기게 된다. 그러므로 순위 선정 기준의 추가는 불가피하다. 그래서 본 논문에서는 표 1과 같은 기준들을 고려하였다. 추가된 기준으로는 감정 극성에 대한 엔트로피, 검색어와 감정 단어의 근접도, 상품평 전체의 감정 단어에 대한 TF가 있다.

표 1 순위 선정 기준 후보군

순위 선정 기준
검색어(Feature)에 대한 TF
감정 극성(Sentiment polarity)에 따른 엔트로피(H)
검색어와 감정단어(Sentiment word)의 근접도
상품평 전체 감정 단어에 대한 TF

감정 극성에 대한 엔트로피는 상품평의 평가가 긍정과 부정 중 어느 한 방향으로 기울었는지 가늠할 수 있는 선정 기준이다. 상품평 d의 엔트로피 H_d는 다음과 같이 계산할 수 있다.

$$H_d = (-\frac{p}{s}\log\frac{p}{s}) + (-\frac{n}{s}\log\frac{n}{s}) \quad (2)$$

식 (2)에서 p는 상품평 내에서 긍정적인 감정 단어의 출현 회수를, n은 부정적인 감정 단어의 출현 회수를, s는 감정 단어 전체의 출현 회수를 의미한다. 엔트로피의 값은 0에서 1사이의 값이 되며, 0에 가까울수록 상품평의 극성이 뚜렷하다는 것을 의미하므로, 사용자의 긍정이나 부정적인 의견이 일관성 있게 표현된 상품평임을 알 수 있다.

검색어와 감정 단어가 근접해 있을 경우, 상품의 특성인 검색어에 대해 사용자가 어떠한 감정을 느꼈는지 더 정확히 알 수 있다. 상품평 내에서 검색어와 감정 단어가 멀리 떨어져 있다면 둘 사이의 직접적인 상관관계는 적을 것이다. 예를 들어 "I am satisfied with the battery life."의 경우에 'satisfied'와 'battery'의 거리차이는 3이다. 내용상으로도 배터리 수명에 만족함을 알 수 있다. 하지만 "This camera is the worst camera I have ever owned. It has many small buttons and only one battery."의 경우 'worst'와 'battery'의 거리차이는 14이며 서로 관련이 없다. 그러므로 검색어의 앞뒤 일정 범위 이내에 감정 단어가 나타날 경우 가중치를 주도록 하였다.

마지막으로 상품평 전체의 감정 단어에 대한 TF로는 상품에 대한 사용자의 감정의 폭을 알아볼 수 있다. 감정이 적은 중립적인 상품평은 상품에 대한 평가보다는 일반적인 설명이 담겨 있는 경우가 많기 때문에, 감정의 폭이 큰 상품평에 더 가중치를 두게 하였다.

4. 실험 결과 및 분석

4.1 실험 방법

3장에서 제시한 상품평 분석 기법의 효율성을 검증하기 위해 실험을 실시하였다. 실험환경의 구축을 위해 프로그램 언어로는 Python 2.5를 사용하였고, 데이터베이스 관리 시스템은 SQLite 3을 사용하였다.

표 1의 네 가지 순위 선정 기준 중, 검색어에 대한 TF를 제일 우선적으로 사용하였다. 나머지 선정 기준들은 표 2와 같이 순서를 조합하여 여섯 번의 실험을 실시하였다. 그리고 일반적 웹 검색엔진과의 결과 비교를 위해, 오픈소스 자바 검색 엔진으로 널리 사용되고 있는 Lucene[7]을 기반으로 한 Nutch[8]에 동일한 데이터를 입력하여 실험하였다. Nutch는 TF*IDF가중치를 이용하여 순위를 선정한다.

표 2 실험별 순위 선정 기준

우선순위	실험 1	실험 2	실험 3	실험 4	실험 5	실험 6
1st	Query TF	Query TF	Query TF	Query TF	Query TF	Query TF
2nd	Entropy	Entropy	SentimentTF	SentimentTF	Distance	Distance
3rd	SentimentTF	Distance	Entropy	Distance	Entropy	SentimentTF
4th	Distance	SentimentTF	Distance	Entropy	SentimentTF	Entropy

Query TF: 검색어(Feature)에 대한 TF
Entropy: 감정 극성(Sentiment polarity)에 따른 엔트로피(H)
Sentiment TF: 상품평 전체의 감정 단어에 대한 TF
Distance: 검색어와 감정 단어(Sentiment word)의 근접도

본 실험은 온라인 쇼핑몰인 Amazon.com의 전자 기기에 대한 400개의 상품평을 대상으로 하였으며, 검색어는 대표적인 특성 중 하나인 'battery'로 선택하였다. 400개의 상품평에 대해서 가장 정확한 검색 결과라고 판단되는 Top 10개와 Top 20개를 수작업으로 결정하였고, 이 결과를 본 논문에서 제시한 검색 기법과 비교하여 정확도를 평가하였다.

4.2 실험 결과

실험 결과는 그림 2, 3과 같다. 우선 그림 2는 Top10개에 대한 정확도 비교 결과이다. 여기서는 Nutch는 70%, 실험군은 평균 73%의 정확도를 보여 큰 차이가 없었다. 다만 실험 1, 2는 다른 실험군에 비해 약간 우수한 결과를 보였다. 실험 1, 2는 검색어에 대한 TF 다음으로 엔트로피를 우선순위로 한 검색 결과로, TF를

제외한 나머지 순위 선정 기준 중에는 엔트로피가 실험 결과에 가장 많은 영향을 미친다고 볼 수 있다.

그림 3은 Top 20개에 대한 실험 결과이다. Top 20개의 상품평에 대해서 Nutch는 60%의 정확도를 보였고, 실험군은 평균 70%의 정확도를 보였다. 따라서 주관적인 의견을 반영하여 상품평을 검색하는 것이 검색어의 출현빈도만을 이용하는 TF*IDF보다 더 정확한 검색 결과를 보여줄 수 있다는 것을 본 실험을 통하여 알 수 있다.

또한, 모든 실험에 대해서 상품평의 엔트로피에 비중을 둔 검색 방법(실험 1, 2)이 가장 좋은 결과를 보였다. 이 결과를 통하여 사용자들은 의견들의 일관성이 뚜렷한 상품평을 긍정과 부정적인 의견이 섞여있는 상품평보다 더 선호한다는 것을 알 수 있다.

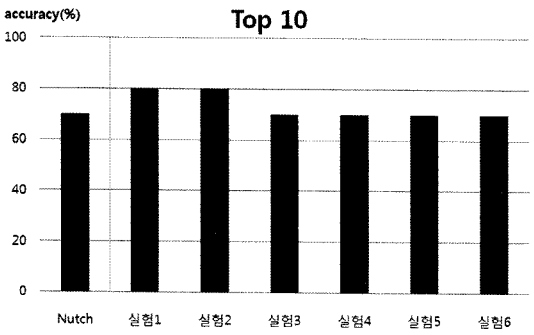

그림 2 실험군과 Nutch의 검색 정확도 비교(Top 10개)

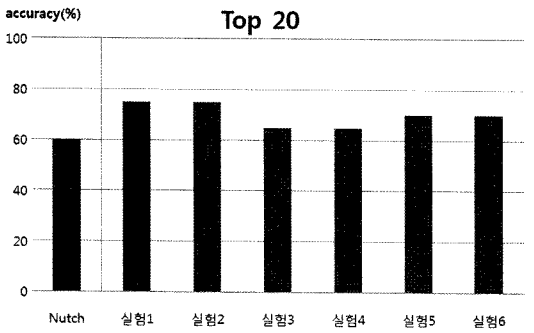

그림 3 실험군과 Nutch의 검색 정확도 비교(Top 20개)

표 3은 사용자가 수작업으로 정한 정답 순위, Nutch 검색 엔진 순위, Top 10개의 문서에 대한 본 실험결과를 나타낸다. 표에서 알파벳은 개별 상품평의 ID를 나타낸다. 상품평 A, B, C, D는 모든 실험군에서 상위에 선정되었다. 따라서 상위의 우선순위는 사용자의 정답 순위와 일치하였다. Nutch의 결과 역시 상위 세 개의 상품평은 사용자의 정답 순위와 일치하였다.

표 3 실험 결과별 Top 10

(※ Bold체 표시는 Nutch와 제안기법이 대조되는 상품평임)

사용자 순위	Nutch	실험1	실험2	실험3	실험4	실험5	실험6
A	A	A	A	A	A	A	A
B	C	B	B	B	B	B	B
C	B	C	C	C	C	C	C
D	**J**	**D**	**D**	N	N	F	F
E	K	G	K	**D**	F	**D**	**D**
G	L	K	G	G	**D**	K	H
L	G	F	F	F	H	H	I
M	P	H	H	H	I	I	K
F	M	E	E	I	G	L	L
Q	Q	Q	Q	M	M	G	G

표 4는 상위 순위에 선정된 상품평 A, B, C의 검색어에 관한 부분을 보여준다. 상품평 A, B, C를 작성한 사용자들은 배터리 지속시간에 만족하고 있음을 알 수 있다. 세 상품평 모두 검색어와 감정 단어가 일정 거리 내에 들어있다는 점도 확인할 수 있다. 하지만 C의 경우 'complain'이라는 부정적인 단어가 들어있지만 'not'에 의해서 긍정적인 문맥으로 변하게 된다. 본 실험에서는 unigram으로 단어의 극성을 판단하였기 때문에 해당 문장을 부정적으로 인식하였음을 확인할 수 있었다.

표 4 주요 상품평의 검색어 관련 내용

	검색어 해당 부분 내용
A	The **battery** life is very very **good**,
	but I always recommend having an extra **battery**($40) on you if you are going abroad with any camera.
	I shot tons of pics in Paris and I think the **battery** lasted a day and a half.
B	**Battery** life is not what is promised but it is still very **good**.
	Of course using the movie will shorten the **battery** life.
C	I can not **complain** about the **battery** life,
	since I was able to take over 150 photo's on one **battery**.

Nutch와 실험군의 Top 10개의 문서중에서 8개는 일치하였지만, 상품평 J와 P는 실험군의 Top 10에 포함되지 않았다. 특히 상품평 J는 Nutch의 결과에서는 네 번째로 순위가 높은 데 비해 실험군에서는 Top 10개의 상품평에 포함되지 못하여 그 순위가 매우 하락했음을 알 수 있었다. 반면 상품평 D는 Nutch의 결과에서는 Top 10개의 상품평에 포함되지 못했지만, 본 실험 결과에서는 상위에 선정되었다.

표 5는 상품평 J, D의 검색어에 관한 실제 내용이다. 상품평 J를 작성한 사용자는 상품에 대해 어느 정도 만

표 5 변동폭이 큰 상품평의 검색어 관련 내용

	검색어 관련 부분 내용
J	Some of the feature were a little tricky but all in all everything was good. The second time the error message redisplayed and the camera turned itself off.
	It turns out that the **battery** was defective.
	I decided to replace the camera.
D	(1) Image quality: superbly clear with vibrant colors and detail. (2) Wide angle lens - The most important reason for buying it and a great advantage for every conceivable shot except one that requires an UZ. ... (8) Sliding door to protect the lens: I did not think I would like this feature, but it is really cool, protects the lens and turns on the camera when you open it.
	(9) **Good battery** life.

족하지만, 배터리에 문제가 있어서 카메라를 교체한다고 이야기한다. 구체적인 특성이 아닌 불량품으로 인한 교체를 이야기하고 있는 것이다. 상품 교환은 발생 가능한 일이지만 보편적으로 일어나는 일은 아니다. 전체적으로 보면 긍정적인 부분과 부정적인 부분이 섞여 있어 상품평의 엔트로피도 높게 나왔기 때문에 실험에서 낮은 평가를 받아서 Top 10에 포함되지 않았다고 볼 수 있다.

반면에 상품평 D는 배터리에 관한 내용이 마지막 부분에 한 번 등장한다. 그 문장을 통해 사용자가 배터리 성능에 대해 만족한다는 점을 알 수 있다. 또한 전체적인 상품평 내에 상품에 대한 감정 단어가 자주 등장하여 사용자의 의견을 구체적으로 표현하고 있다. 상품에 대해서 일관성 있게 매우 긍정적으로 평가하고 있는 것을 알 수 있다. 그렇기 때문에 상품평 D는 Nutch의 결과와는 달리 실험 결과에서 매우 높은 순위로 선정되었다.

5. 결론 및 향후 연구

상품평은 다른 웹 문서와는 달리 감정적이며 주관적일 수밖에 없다. 상품평에 포함된 상품의 특성과 의견의 양, 그리고 내용을 정확히 추출하는 것이 중요하다. 본 논문에서는 이러한 상품평의 특징에 어울릴 수 있도록 상품평 내에 주관적인 의견의 포함 여부 및 그 정도 등을 고려한 순위 선정 기준을 마련하여 실험하고 그 결과를 분석해보았다. 실험 결과 의견이 분명히 드러나거나 정리가 잘 된 상품평이 높은 순위에 선정됨을 확인하였다. 또한 부분적으로 일반적인 검색 엔진의 검색 결과보다 상품평 검색에 대한 사용자의 의도와 일치하는 것을 우선적으로 검색할 가능성을 더욱 높일 수 있다는 결론을 내릴 수 있다.

향후 연구 과제로는 multigram단위로 문장을 분석하는 것이다. 현재 unigram단위로 문장을 분석하였기에 문장의 긍정과 부정에 대한 극성을 반대로 인식하는 경우가 있었다. 단어만이 아닌 절과 구도 반영하여 긍정과 부정의 판단을 할 수 있도록 해야 할 것이다. 그리고 순위 선정 기준에 대한 우선순위를 두는 현재의 실험 방법 외에도, 모든 선정 기준들이 순위 선정에 영향을 미칠 수 있도록 적절한 비율을 부여하는 방안을 추진할 것이다.

참 고 문 헌

[1] Xiaowen Ding, Bing Liu, "The Utility of Linguistic Rules in Opinion Mining," *Proc. of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval 2007*, pp.811-812, 2007.

[2] Pavel Smrž, "Using WordNet for Opinion Mining," *Proc. of the International WordNet Conference 2006*, pp.333-335, 2006.

[3] Qingliang Miao, Qiudan Li, Ruwei Dai, "AMAZING: A Sentiment Mining and Retrieval System," *Expert Systems with Applications*, vol.36, no.3, pp.7192-7198, 2009.

[4] Andrea Esuli, Fabrizio Sebastiani, "PageRanking WordNet Synsets: An Application to Opinion Mining," *Proc. of the Annual Meeting of the Association for Computational Linguistics 2007*, pp.424-431, 2007.

[5] Jung-Yeon Yang, Jaeseok Myung, Sang-goo Lee, "A Sentiment Classification Method Using Context Information in Product Review Summarization," *Journal of KIISE : Databases*, vol.36, no.4, pp.254-262, 2009. (in Korean)

[6] Bo Pang, Lillian Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol.2, no.1-2, pp.1-135, 2008.

[7] http://lucene.apache.org/

[8] http://lucene.apache.org/nutch/