

# 일반연구자지원사업 최종(결과)보고서

							양식A101	
① 부처사업명(대)	기초연구사업				보안등급(보안, 일반)		일반	
② 사 업 명(중)	일반연구자지원사업				공개가능여부(공개, 비공개)		공개	
③ 세부사업명(소)	여성과학자							
④ 과제성격(기초, 응용, 개발)	기초		④-1 실용화 대상여부(실용화, 비실용화)					
⑤ 과 제 명	국 문	감성사전을 이용한 트위터 감성 인식						
	영 문	Twitter sentiment analysis using sentiment lexicon						
⑥ 주관연구기관	충남대학교							
⑦ 협동연구기관								
⑧ 주관연구책임자	성 명	이 공 주			직급(직위)	부교수		
	소속부서	정보통신공학과			전 공	전산		
⑨ 연구개발비 및 참여연구원수 (단위: 천원, M·Y)								
년 도	정부출연금 (A)	기업체부담금			정부외 출연금 (B)	상대국 부담금 (F)	합계 G=(A+B+E)	참여 연구원수
		현금 (C)	현물 (D)	소계 E=(C+D)				
1차년도	38,100			0			38,100	8
2차년도	38,100			0			38,100	5
3차년도				0			0	
4차년도				0			0	
5차년도				0			0	
합계	76,200	0	0	0	0	0	76,200	13
⑩ 총연구기간		2011. 05. 01 ~ 2012. 04. 30 (24개월)						
⑪ 다년도협약연구기간		기재하지 않음						
⑫ 당해연도연구기간		2012. 05. 01 ~ 2013. 04. 30(12개월)						
⑬ 참여기업	중소기업수		대기업수		기타		계	
							0	
⑭ 국제공동연구	상대국연구기관수		상대국연구개발비		상대국연구책임자수			
<p>관계 규정과 모든 지시사항을 준수하면서 국가연구개발사업에 따라 수행 중인 연구개발과제의 최종보고서를 붙임과 같이 제출 합니다.</p> <p style="text-align: center;">2013 년 7 월 9 일</p> <p style="text-align: center;">주관연구책임자 : 이 공 주</p> <p style="text-align: center;">주관연구기관장 : 충남대학교 산학협력단장</p> <p style="text-align: center;"><b>교 육 과 학 기 술 부 장 관 귀 하</b></p>								

※ 전자접수이므로 주관연구책임자 및 주관연구기관장 서명(인, 직인)은 생략

**【주요항목 작성요령】**

- ①, 부처사업명(대), ② 사업명(중), 보안등급(일반), 공개가능여부(공개)는 수정하지 않음
- ③ 세부사업명(소)은 **기본연구(유형 I), 기본연구(개인), 기본연구(모험), 기본연구(커리어), 신진연구, 신진연구(연구비), 신진연구(연구장비), 여성과학자, 지역대학우수과학자 중 택일**
- ④ 과제성격 및 ④-1 실용화 대상여부는 수정하지 않음
- ⑤ 과제명은 당초 연구과제명(과제명 변경을 재단에서 승인받은 경우는 승인된 과제명)을 기재함
- ⑥ 주관연구기관은 한국업적통합정보(KRI)에 입력된 기관명과 동일해야 하며 약어를 사용하지 않음(서울대→서울대학교)
- ⑦ 협동연구기관은 본 사업과 관련이 없으므로 기재하지 않음
- ⑧ 연구책임자의 인적사항 기재
- ⑨ 정부출연금은 전체연구기간동안 기 지원받은 연도별 총 연구비(간접비 포함)를 기재하며, 참여연구원은 연구책임자를 제외한 참여 연구인력(연구보조원 포함) 인원수를 기재함
- ⑩ 총연구기간은 연구시작일부터 연구종료일까지의 총연구기간 및 개월 수를 기재함  
⇒ 아래의 연구기간 적용안내 참조
- ⑪,⑫,⑬,⑭는 기재하지 않음

**「일반연구자지원사업 연구기간 적용 안내」**

선정연도	연구기간	총 연구기간	다년도 협약기간	당해연도 연구기간	비고
2010년	36개월	2010.05.01~2013.04.30	기재하지 않음	기입력(수정하지 않음)	
2011년	24개월	2011.05.01~2013.04.30	기재하지 않음	기입력(수정하지 않음)	
2012년	12개월	2012.05.01~2013.04.30	기재하지 않음	기입력(수정하지 않음)	

## < 목 차 >

### I. 연구결과 요약문

### II. 연구내용 및 결과

1. 연구과제의 개요 ..... 5
2. 국내·외 기술개발 현황 ..... 5
3. 연구수행 내용 및 결과 ..... 5
4. 목표 달성도 및 관련 분야에의 기여도 ..... 8
5. 연구결과의 활용 계획 ..... 8
6. 연구과정에서 수집한 해외 과학기술 정보 ..... 8

### III. 연구성과

## I. 연구결과 요약문

본 연구과제는 트위터로 대표되는 소셜 네트워크 서비스(SNS)에 나타난 대중의 여론 분석이 주요 목적이다. 빅데이터의 대표적인 예인 SNS서비스의 사용자 개개인의 의견이 담긴 글을 분석함으로써 인물, 제품, 또는 사건 등에 대한 감성을 분석하고, 이를 통해 개개인의 의견을 모아 대중이 어떤 대상에 대해 가지고 있는 여론을 도출해 낼 수 있다.

### 1. 대량의 트위터 수집, 분류, 저장, 검색 시스템 연구

이를 위해 트위터로부터 대량의 정보를 수집, 분류, 저장, 검색에 필요한 시스템의 연구를 진행하였다. 본 연구과제가 수행하고 있는 연구를 통해 일반 대중들은 자신이 선택하고자 하는 제품, 정치인 등에 대한 다른 대중들의 선호를 파악 하여 선택에 반영할 수 있고, 기업, 정치인들 입장에서는 자신들에 대한 대중의 피드백을 확인 할 수 있게 되므로 학술적으로 연구 가치가 있고 사회적으로도 이익이 될 수 있다. 이 시스템의 유용성 검증을 위하여 2012년 대선의 후보군에 대해 적용한 프로토타입 시스템을 제작하였고, 각 컴포넌트 별로 수정과 교체를 하며 시스템의 완성도를 높이고 있다.

### 2. 트렌드 토픽과 감성사전을 이용한 SNS사용자의 여론 파악 연구

인물, 제품만이 아닌 사회에서 일어나고 있는 사건, 사고(트렌드 토픽)에 대한 대중의 여론 파악을 위한 연구도 진행하였다. 여론 파악의 방법으로 개개인의 트위터를 감성사전을 이용하여 감성을 분석하는 방법에 대하여 연구 하였다. 트렌드 토픽에 대한 트위터상의 여론을 알아보기 위하여 인터넷 신문으로부터 트렌드 토픽에 관한 키워드를 추출하고, 추출된 트렌드 토픽을 이용하여 트위터의 수집, 분석을 수행하는 과정에 대하여 연구하였다. 이러한 과정을 위하여 인터넷 신문 기사를 수집하고, 수집된 신문 기사를 클러스터링 알고리즘을 이용하여 분류하였다. 분류된 인터넷 신문 기사들을 이용하여 각 클러스터를 구성하는 신문 기사들을 대표할 수 있는 키워드를 추출 하고 대표성을 평가 하였다. 인터넷 신문 기사를 통해 수집된 토픽에 대한 트위터 데이터를 수집하고 이에 대하여 감성사전을 이용한 감성 분석 작업을 수행하였다.

### 3. 향상된 토픽 추출과 감성 분석 방법에 대한 연구

인터넷 신문기사를 활용한 향상된 트렌드 토픽 추출 방법과 추출된 토픽에 대한 대중의 관심도, 확률모델을 통한 감성분석 방법에 대한 연구도 진행하였다. 신문기사의 모음으로부터 사회 이슈를 찾아내고, 그 이슈에 대한 트위터상의 감성을 파악할 수 있는 시스템에 대하여 연구하였다. 그리고 실험을 통해 그 가능성을 보였다. 이를 통해 해당 주제에 대한 대중의 긍정, 부정의 감성을 파악할 수 있게 되어 주제에 대한 여론의 향방을 살펴 볼 수 있게 된다. 이러한 기술과 토픽탐지 및 추적 기술을 결합하여 정치, 연예, 증시관련 이슈의 변화와 그에 따른 대중들의 반응을 추적하여 활용할 수 있을 것이다. 연구를 위해 선정된 트렌드 토픽의 화제성과 트윗의 감성분석 결과에 대한 상관관계 실험에서 65.97%의 결과를 보였다. 이는 앞으로 트위터 감성분석의 성능을 높이고, 신문기사 군집화와 토픽 추출의 성능을 향상시킴으로써 높여나갈 수 있으리라 생각하고 계속 연구해야 할 주제이다. 본 연구에서 수행한 일련의 작업은 각 단계마다 다양한 접근방법과 연구주제를 가지고 있다. 신문기사 군집화와 토픽 추출, 트위터 분석 등 다양한 연구 분야와 주제가 혼합 되어 있어 각각의 주제에 대하여 심화된 연구를 통해 본 시스템의 효용성을 높일 수 있을 것이라 판단하고 이는 향후 연구과제로 남아 있다.

## II. 연구내용 및 결과

- ◎ 1. 연구과제의 개요 ~ 6 기타사항을 항목에 따라 작성함
- ◎ 제목 14point, 소제목 12point, 본문내용은 10point로 작성하며, 줄 간 간격은 조정 가능함
- ◎ 내용 작성과 관련한 설명내용(청색 박스로 표시된 부분)은 내용 작성 시 제거하고 기술함

※ II. 연구내용 및 결과의 작성분량은 3페이지 내외로 핵심내용 중심으로 간결하게 작성함

### 1. 연구과제의 개요

#### 연구의 목적

이전까지의 여론 조사는 주로 오프라인 상에서 직접 대면을 통한 설문이나 대상에게 전화를 걸어 설문 조사하는 방법이 주류였다. 하지만 최근에 트위터의 발달로 일반 대중들이 자신의 생각이나 자신이 접한 이슈에 대하여 자신의 트위터에 올려 여러 사람들과 소통하는 경우가 많아졌다. 많은 사람들이 많은 주제에 대하여 저마다의 생각을 작성하고, 이를 공유하고 토론함으로써 트위터는 이미 기성 언론이나 인터넷 게시판이 가지던 여론 형성 기능과 맞먹는 정도의 여론의 장이 되어가고 있다. 그렇기 때문에 어떠한 토픽에 대하여 대중들의 생각을 알아보기 위한 장소로 트위터가 큰 위력을 발휘할 수 있다. 이와 같은 트위터의 장점을 이용하기 위해 트위터에 올라온 글(트윗)을 분석하는 방법에 대한 연구가 본 연구의 목적이다. 이를 위하여 감성사전을 이용한 트윗의 감성 분석, 확률 모델을 이용한 감성분석, 화제가 되는 단어의 추출과 분석방법 등에 대한 연구를 수행하였다.

### 2. 국내외 기술개발 현황

기존의 트위터 감성 분석 연구의 대부분은 해당 트윗의 주제에 대한 고려 없이 트위터에 나타난 긍, 부정의 감성추출에 중점을 두었다. 이와 다르게 본 연구에서는 트윗에 나타난 주제를 파악하고, 이 주제에 대한 트윗의 감성을 판별하였다.. 특정 시기의 신문기사 그룹에서 중요하게 다루어지는 단어들은 그 시기에 많은 사람들이 언급하는 화제가 되는 키워드일 것이다. 그렇기 때문에 화제가 되는 단어들(토픽 키워드)을 포함하고 있는 트윗이 있다면, 그 트윗은 해당 화제에 대하여 언급하고 있는 트윗일 것이라고 가정하였다. 본 연구에서는 이와 같은 가정과 함께 인터넷 신문기사와 트위터가 가지는 특징을 이용하여 화제가 되는 최근 동향에 대한 일반 대중의 여론을 분석하는 방법을 연구하였다.

### 3. 연구수행 내용 및 결과

본 연구 내용의 전체적인 개요는 아래(그림 1)과 같다. 주요 단계에 대하여 설명한다.

#### 1. 주제별 군집화

진행과정 1에서 수집된 신문기사는 다양한 주제를 가지고 있다. 서로 유사한 주제를 다룬 신문기사를 군집화하여 각 군집을 대표하는 키워드를 추출하여 트렌드 토픽으로 활용한다. 군집의 크기를 이용하여 토픽의 중요성을 반영할 수 있다. 주제별 군집화에는 대표적인 군집화 알고리즘인 k-means 알고리즘을 이용하였으며, 더 나은 신문기사 군집화 방안에 대한 연구를 진행하고 있다.

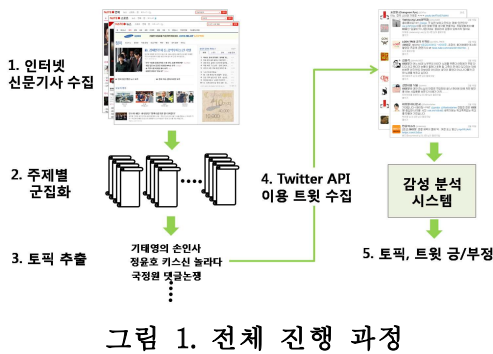


그림 1. 전체 진행 과정

["지나", VV], ["지자을", NNG+XSN]	-0.48
["노", ETM], ["바삼", NNG]	0.572
["정권", NNG], ["을", JKO]	0.7
["의", JKG], ["보여주", VV]	
["실패", NNG], ["는", ETM]	
["에", JKB], ["그", NP]	
["책임", NNG], ["를", JKO]	
["이", JKS], ["적극적", NNG+XSN]	
["있", VX], ["으로", JKB]	
["기", ETN], ["지지하", NNG+XSV]	
["때문", NN8], ["지", E]	
["에", JKB], ["못하", VX]	
[" ", SP], ["다", E]	
["최근", NNG], [".", SF]	
["무섭", VA]	
["노", ETM]	

["실패", NNG]  
["에", JKB]  
["책임", NNG]  
["있", VX]

앞	한
	못
뒤	지 않다
	지 못하다
	아니다
	못하다

그림 2. 감성사전을 이용한 감성분석 예시

## 2. 토픽 키워드 추출

군집화된 신문기사에서 토픽 키워드 방법은 크게 2가지 접근 방법에 대한 실험을 수행하였다.

### a) tf · idf\_top3, PMI\_top3, pmi-tf · idf\_top3

정보검색 분야에서 문서에 나타난 단어의 중요성을 나타내는 점수인 tf · idf(문서 내 단어 출현 빈도 · (1/단어가 출현한 문서의 빈도)점수와 확률론에 기반한 두단어의 동시 출현 빈도(pmi)점수를 활용한 키워드 추출 방법을 연구하였다. 이 방법을 통한 키워드 추출단계(표. 1)와 결과(표. 2)는 아래와 같다.

1) PMI-tf · idf_top3의 1번을 구성하는 w1, w2중 w1이 tf · idf_top3에 속한다면 topic에 w1과 w2를 포함시킨다.
1.1) 1을 만족한 경우 PMI-tf · idf_top3의 2번을 구성하는 w1, w2 중 w1이 tf · idf_top3에 속한다면 topic에 w1과 w2를 포함시킨다.
1.1.1) 1.1을 만족한 경우 PMI-tf · idf_top3의 3번을 구성하는 w1, w2 중 w1이 tf · idf_top3에 속한다면 topic에 w1과 w2를 포함시킨다.
2) PMI_top3의 1번을 구성하는 w1, w2중 w1이 tf · idf_top3에 속한다면 topic에 w1과 w2를 포함시킨다.
2.1) 2을 만족한 경우 PMI_top3의 2번을 구성하는 w1, w2 중 w1이 tf · idf_top3에 속한다면 topic에 w1과 w2를 포함시킨다.
2.1.1) 2.1을 만족한 경우 PMI_top3의 3번을 구성하는 w1, w2 중 w1이 tf · idf_top3에 속한다면 topic에 w1과 w2를 포함시킨다.
3) 1,2의 조건을 만족 시키지 못한 경우 tf · idf_top3를 topic으로 인정한다.

표 1. 트렌드 키워드 추출 단계

클러스터	tf-idf_top3	pmi_top3	pmi-tf-idf_top3	토픽
2012-11-08 정치 클러스터 25번	일자리청 신설 실무적	제주, 직원들 정책, 발표회 개최, 발표회	일자리청, 신설 실무적, 일자리청 실무적, 신설	일자리청 신설
2012-11-12 연예 클러스터 62번	조혜련 이혼 힐링캠프	광격택, 오리새끼 오리새끼, 원더우먼 조혜련의 사연들	조혜련 이혼 힐링 서프라이즈 예상 서프라이즈	조혜련 이혼

표 2. 트렌드 토픽 추출 예시

### b) 군집 내 신문기사 수와 단어 출현 빈도 활용

군집화 결과 신문기사 개수가 2개인 군집의 경우 출현 빈도가 가장 높은 하나의 단어를 토픽 키워드로 선정하였고, 군집 내의 기사 개수가 3개인 경우 출현 빈도 상위 3개의 단어를, 신문기사 4개 이상인 경우 상위 5개의 단어를 군집의 토픽 키워드로 선정하였다. 같은 출현 빈도의 경우 개체명과 명사를 우선순위에 두었다. 추출한 토픽 키워드가 신문기사 군집에 대한 주제 키워드로서 적합한지에 대하여 평가하였다. 평가 방법은 2명의 평가자가 신문 군집의 기사를 보고 떠오르는 키워드 5개를 작성하게 하고, 프로그램을 통해 추출된 키워드와 비교하여 정보검색분야에서 평가 척도 중 하나로 사용되는 R-Precision을 통해 평가하였다. 그 결과는 아래(표 4)와 같다.

	Tester 1	Tester 2
R-Precision	0.866	0.908

표 4. 트렌드 키워드 평가 결과

권위적이지 않다	+2
기대에서 벗어나다	-1
길들여지지 않다	0

표 5. 감성사전 예

### 3. 토픽트윗 긍/부정 검출

수집된 트윗을 분석하여 트윗이 토픽 키워드에 대하여 나타내고 있는 감성을 분석한다.

분석 방법에는 크게 a) 감성사전을 이용하여 감성을 분석하는 방법과 b) 학습데이터를 이용한 확률모델을 통한 감성 분석 방법으로 나눌 수 있다.

#### a) 감성사전을 이용한 분석 방법

감성 사전은 대상에 대한 감성을 표현하는 표현들과 그 감성의 극성값을 모아놓은 사전이다. 그 예는 (표 5)와 같다. 본 연구과제를 통하여 총 4,728개의 감성 표현을 수집하고 이를 실험에 활용하였다.

감성사전을 이용한 감성값 획득 방법은 다음과 같다.

- a) 감성을 판별하려는 트윗을 분석 대상이 되는 토픽에 대한 트윗으로 가정한다.
- b) 형태소 분석된 감성 표현의 첫 형태소가 대상 트윗의 형태소 분석 결과에 포함되어 있는지 확인한다.
- c) 포함되어 있다면, 대상 트윗에서 감성 표현의 형태소가 검출된 위치 다음부터 나머지 감성표현의 형태소가 포함되어 있는지를 확인한다.
- d) c)에서 감성 표현의 포함이 확인된 트윗의 경우, 부정어 표현이 대상 트윗에 포함되어 있는지 확인한다. c)에서 모여진 감성 표현들에서, 각 감성 표현이 검출된 위치의 앞부분에 “안”, “못” 과 같이 대상의 앞에서 부정을 표현해주는 표현이 있다면 해당 표현이 부여하는 감성 값을 반전시켜준다.

표 6 감성사전을 이용한 감성값 판별 방법

이를 도식화 하면 (그림 2)와 같다.

#### b) 학습데이터를 이용한 확률모델을 통한 감성 분석 방법

평가자에 의해 극성값이 할당된 트윗 중 감성 없음 트윗을 제외한 2,080개의 트윗(긍정 1,277개, 부정 803개)을 이용하여 감성값 할당 실험을 수행하였다. 각 자질 조합의 실험 결과를 (표 7)에 나타내었다.

감성분석 실험에 사용된 트윗을 이용하여 토픽 키워드와 트윗간의 관계를 평가하였다. 평가 방법은 토픽 키워드와 그로부터 추출된 트윗에 대하여 1)트윗이 토픽 키워드에 대하여 언급하고 있는지, 2)감성분석 결과가 토픽 키워드에 대한 감성과 일치하는지에 대하여 평가하였다.

평가 결과를 (표 8)에 나타내었다.

Feature List	Accuracy
baseline	62.8%
unigram(f1)	72.64%
unigram(f1)+bigram(f2)	73.29%
Twitter 관련 자질(TWF)	60.92%
f1+f2+TWF	74.46%

표 7. 확률 모델을 이용한 감성분석 실험결과

topic-tweet relation	tweet-senti relation	topic-tweet-senti relation
88.42%	74.46%	65.97%

표 8. 트윗-토픽키워드간 감성 분석 결과

(표 8)의 topic-tweet relation은 트윗의 주제가 토픽 키워드이며 해당 트윗에 나타난 감성이 토픽과 관련된 경우의 비율을 나타낸다. tweet-senti relation은 트윗에서 토픽을 고려하지 않는 감성 분석의 결과로 (표 7)의 실험 중에서 가장 좋은 결과를 나타낸다. topic-tweet-senti relation은 트윗의 주제가 토픽이며, 감성분석 실험 결과의 결과와 실제 트윗에서 나타난 토픽에 대한 감성이 일치하는 비율을 나타낸 것이다. 실험 결과에서 토픽 키워드로 수집된 트윗 중 그 토픽에 대하여 말하고 있는 트윗의 비율이 88.42%로 나타나 앞서 가정하였던 토픽 키워드와 그로부터 수집된 트윗 간의 관계에 대한 가정의 타당성을 보이고 있다. 이는 트윗의 감성 분석을 수행 할 때 트윗이 화제성이 있는 키워드로 수집된 트윗이라면 구문분석단계 등을 거치지 않아도 해당 트윗의 감성을 주제 키워드에 대한 것으로 간주할 수 있다는 결론을 이끌어 낼 수 있다.

## 4. 목표 달성도 및 관련 분야에의 기여도

본 연구에서는 신문기사의 모음으로부터 사회 이슈를 찾아내고, 그 이슈에 대한 트위터상의 감성을 파악할 수 있는 시스템에 대하여 연구하였다. 그리고 실험을 통해 그 가능성을 보였다. 이를 통해 해당 주제에 대한 대중의 긍정, 부정의 감성을 파악할 수 있게 되어 주제에 대한 여론의 향방을 살펴 볼 수 있게 된다. 이러한 기술과 토픽탐지 및 추적 기술을 결합하여 정치, 연예, 증시관련 이슈의 변화와 그에 따른 대중들의 반응을 추적하여 활용할 수 있을 것이다. 본 연구에서 가정한 화제성과 트윗의 감성분석 결과에 대한 상관관계 실험에서 65.97%의 결과를 보였다. 이는 앞으로 트위터 감성분석의 성능을 높이고, 신문기사 군집화와 토픽 추출의 성능을 향상시킴으로써 높여나갈 수 있으리라 생각하고 계속 연구해야 할 주제이다.

## 5. 연구결과의 활용계획

이전까지의 여론 조사는 주로 오프라인 상에서 직접 대면을 통한 설문이나 대상에게 전화를 걸어 설문 조사하는 방법이 주류였다. 하지만 최근에 트위터의 발달로 일반 대중들이 자신의 생각이나 자신이 접한 이슈에 대하여 자신의 트위터에 올려 여러 사람들과 소통하는 경우가 많아졌다. 많은 사람들이 많은 주제에 대하여 저마다의 생각을 작성하고, 이를 공유하고 토론함으로써 트위터는 이미 기성 언론이나 인터넷 게시판이 가지던 여론 형성 기능과 맞먹는 정도의 여론의 장이 되어가고 있다. 그렇기 때문에 어떠한 토픽에 대하여 대중들의 생각을 알아보기 위한 장소로 트위터가 큰 위력을 발휘할 수 있다. 본 연구에서 수행한 연구를 통해 일반 대중들은 자신이 선택하고자 하는 제품, 정치인등에 대한 다른 대중들의 선호를 파악 하여 선택에 반영할 수 있고, 기업, 정치인들 입장에서는 자신들에 대한 대중의 피드백을 확인 할 수 있게 되므로 학술적으로 연구 가치가 있고 사회적으로도 이익이 되는 시스템이 될 것이라 생각한다.

## 6. 연구과정에서 수집한 해외과학기술정보

(Akshat et al., 2012)는 학습 데이터 안의 unigram들의 긍정, 부정 트윗에서 나타난 발생빈도 간의 차를 이용하여 긍정, 부정을 분류하는 1)Baseline과 2)긍정, 부정, “!” 문자의 전처리, 3)각 단어들의 어근 추출, 4)불용어 제거, 5)철자 교정, 6)미리 정의된 감성 표현들을 이용하는 감성어휘 자질, 7)명사 제거, 8)unigram의 긍정, 부정에서의 출현 빈도 간의 차의 크기에 따라 긍, 부정의 점수에 차등을 두는 *Popularity Score* 방법을 가지고 각 단계를 조합하는 실험을 수행하였다[4]. Stanford Dataset을 이용한 실험에서 1)Baseline방법은 78.8%의 정확도를 보였고, 각 단계의 모든 조합 결과 87.2%의 정확도를 보였다. Meja Dataset에서는 Baseline 방법이 77.1%, 모든 조합에서 88.1%의 정확도를 보였다.

(Jiang et al., 2011)에서는 쿼리를 포함한 트윗을 수집하고, 감성 분석과정에서 검출된 감성 표현이 해당 쿼리와 의존관계를 가지고 있는지 확인한 후 의존 관계가 있을 경우 감성을 부여하도록 한다[5]. 이 논문의 연구에서는 교사학습 방법인 SVM 알고리즘을 사용하면서, 자질의 분류를 검출 대상에 독립적인 자질과 검출 대상과 비독립적인 자질로 구분하였다. 실험 결과, 검출 대상에 독립적인 자질과 비독립적인 자질 모두 사용하였을 경우 85.6%의 정확도를 보였다.

## 참고자료

[1] 이경호, 이공주, “트위터를 이용한 인물 평판 분석시스템”, 한글 및 한국어 정보처리 학술대회, 2012

[2] 이경호, “신문기사로부터 추출한 최근동향에 대한 트위터 감성분석”, 석사 학위 논문, 2012

[3] 이경호, 이공주, “신문기사로부터 추출한 최근동향에 대한 트위터 감성분석”, 소프트웨어 및 데이터 공학, 한국정보처리학회, 2013

[4] A. Bakliwal, P. Arora, S. Madhappan, N. Kapre, M. Singh, and V. Varma, "Mining sentiments from Tweets," in *WASSA 2012*, pp. 11-18, 2012

[5] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp.151-160, 2011



### Ⅲ. 연구성과

전문학술지 논문게재 성과정보											
과제번호	게재연월	논문제목	총저자명	출처	학술지명	권(호)	학술지구분	sci여부	impact Factor	국제공동연구논문	기여도
2011-007945	201305	Compositional rules of Korean auxiliary predicates for sentiment analysis	K o n g Joo Lee	직접입력	한국마린엔지니어링학회지	37(3)	국내	SCI미등재		아니오	100