

텍스트의 정서 단어 추출을 통한 문학 작품의 정서 분석*

Analyzing Emotions in Literature by Extracting Emotion Terms

함준석** · 이신영** · 고일주**†

Jun-Seok Ham** · Shin-Young Rhee** · Il-Ju Ko**†

송실대학교 미디어학과**

Department of Media, Soongsil University**

Abstract

We define a 'dominant emotion' as acting dominantly for unit time, and propose methodology to extract dominant emotion in a literature automatically. Due to the nature of the Korean language, it is able to be changed or reversed owns meanings as desinence. But it might be possible to extract a dominant emotion in a text has a small quantity like a fiction or an essay. A process to extract a dominant emotion in a literature is as follows. At first, extract morphemes in a whole text. And dispart words having emotional meaning as matching emotion terms database. Map disparted terms to a affective circumplex model and matching it with basic emotion. Finally, analyze dominant emotion according to matched basic emotion. And we adjust our methodology to two literature; modern fiction 'A lucky day' by Jingeon, Hyun and essay 'An old man who shave a bat' by Woyoung, Yun. As a result, it was possible to grasp flows of dominant emotion.

Keywords : Emotion Terms, Literature, Extracting Emotion Terms, Analyzing Emotion

요약

본 논문에서는 단위 시간 동안 주로 작용하는 정서를 '지배적 정서(dominant emotion)'라고 정의하고, 문학작품의 지배적 정서 흐름을 자동적으로 추출하기 위한 방법론을 제시한다. 한국어는 언어 구조적 특성상 접미어에 따라 의미가 역전되거나 달라질 수 있다. 하지만 소설이나 수필 같이 일정 이상의 분량을 가진 텍스트에서 정서 단어를 추출한다면 어느 정도 추출이 잘못되어도 지배적 정서 흐름을 판단하는 것이 가능할 것이다. 문학작품에서 지배적 정서를 추출하기 위한 절차는 다음과 같다. 먼저 문학작품의 전체 텍스트에서 형태소를 분석하여 형태소 단위의 단어를 추출한다. 추출된 단어를 정서 단어 데이터베이스와 매칭하여 정서적 의미를 담고 있는 단어를 분리해 낸다. 분리된 단어들을 정서 모델에 사상하여 해당 단어가 갖고 있는 정서를 도출한다. 도출된 정서 단어들을 통해 지배적 정서를 분석한다. 제안한 방법론에 따라 현진건의 현대소설 '운수 좋은 날'과 윤오영의 수필 '방망이 깎던 노인'을 분석한 결과, 지배적 정서의 흐름을 파악할 수 있었다.

주제어 : 정서 단어, 문학 작품, 정서 단어 추출, 정서 분석

* 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2010-0016540).

† 교신저자 : 고일주 (송실대학교 미디어학과)

E-mail : andy@ssu.ac.kr

TEL : 02-820-0719

FAX : 02-822-3622

1. 서론

문학 작품들은 각각 특정 정서적 흐름을 가지고 있다. 예를 들어 현진건의 현대소설 ‘운수 좋은 날’의 경우 아내에 대한 걱정, 손님이 없는 불안에서 시작해서 운수 좋게 손님이 많은 기쁨을 거쳐 아내의 죽음으로 인한 슬픔으로 끝나는 흐름을 가지고 있다.

독자들은 독서 성향이나 정서적 상태에 따라 특정한 정서적 흐름을 선호한다. 어떤 독자는 고난 끝에 행복한 결말이 있는 흐름을 선호하고, 어떤 독자는 행복한 과정 끝에 슬픈 결말이 있는 흐름을 선호한다. 또한 이러한 선호성향은 즐거운 상태, 우울한 상태 같은 정서적 상태에 따라 바뀔 수 있다.

본 논문에서는 단위 시간 동안 주로 작용하는 정서를 ‘지배적 정서(dominant emotion)’라고 정의하고, 문학작품의 지배적 정서 흐름을 자동적으로 추출하기 위한 방법론을 제시한다. 만약 문학 작품에서 지배적 정서흐름을 자동적으로 분석할 수 있다면, 유사한 정서적 흐름을 가진 문학 작품별 분류를 통해 효과적인 검색과 추천 시스템을 구현하는 것이 가능하다. 독자들은 이미 읽었거나 알고 있는 문학 작품을 통해 정서적으로 유사한 내용을 가진 문학 작품을 검색하고 추천받을 수 있게 된다.

한국어는 언어 구조적 특성상 접미어에 따라 의미가 역전되거나 달라질 수 있다. 따라서 한국어로 된 문장에서 정서적 의미를 담고 있는 단어의 어간만을 추출하여 그 문장이 어떤 정서를 담고 있는지 판단하면 잘못된 결과가 나올 수 있다. 하지만 소설이나 수필 같이 일정 이상의 분량을 가진 텍스트에서 정서 단어를 추출한다면 어느 정도 잘못된 결과가 나오더라도 지배적 정서 흐름을 판단하는 것이 가능할 것이다.

따라서 본 논문에서는 문학 작품에서 정서 단어를 추출하여 단어가 가지고 있는 정서를 시간의 흐름에 따라 나열하면 지배적 정서 흐름을 알 수 있다고 가정한다. 이에 따라 정서 단어 추출 방법과 지배적 정서 흐름 분석 방법을 제시하고 실제 적용을 통해 결과를 분석한다.

문학 작품에서 지배적 정서를 추출하기 위한 절차는 다음과 같다. 먼저 문학 작품의 전체 텍스트에서 형태소를 분석하여 형태소 단위의 단어를 추출한다. 추출된 단어를 정서 단어 데이터베이스와 매칭하여 정서적 의미를 담고 있는 단어를 분리해 낸다. 분리된

단어들을 정서 모델에 사상하여 해당 단어가 갖고 있는 정서를 도출한다. 도출된 정서 단어들을 통해 지배적 정서를 분석한다.

본 논문은 다음과 같이 구성된다. 2장에서는 인공지능과 심리학 분야에서 관련된 연구를 기술하고 3장에서는 정서 단어를 추출하기 위해 사용된 형태소 분석기, 정서 단어 데이터베이스, 정서 모델을 소개하고 이를 통해 정서 단어를 추출하는 방법을 제시한다. 4장에서는 제시한 방법에 따라 실제로 문학 작품 두 편을 분석하고 평가한다. 5장에서는 평가에 대한 결론과 논의를 다룬다.

2. 관련연구

텍스트에서 정서를 가진 단어를 추출하여 텍스트의 정서 상태를 분석하려는 시도는 크게 인공지능 분야와 심리학 분야에서 진행되어 왔다.

인공지능 분야에서는 채팅 텍스트로부터 문장을 분석하여 10가지의 정서 상태로 분류하고 은닉 마르코프 모델(HMM: Hidden Markov Model)을 사용하여 정서 상태의 천이를 모델링 한 연구가 있었다(문현구와 장병탁, 2001). 이 연구에서는 단어뿐 만 아니라 이모티콘과 같이, 단어는 아니지만 정서를 표현하는 데이터를 사용하였다. 그러나 정의된 10가지의 정서 상태가 심리학적 근거가 없이 연구자가 임의에 의해 설정되었고 10가지 정서 상태에 해당하는 단어나 이모티콘도 연구자 임의에 의해 분류되었다는 한계가 있다.

또 다른 연구에서는 단어의 의미를 분석하여 긍정, 부정 자질 등 단어의 정서가(극성)를 추출하여 지지 벡터 기계(SVM: Support Vector Machine)를 사용하여 텍스트 문서를 분류하였다(황재원과 고영중, 2008). 한국어에는 의미 정보를 제공하는 시소러스가 없기 때문에 한국어 단어를 영어 사전을 사용하여 변환한 후 영어 시소러스를 사용하여 정서가를 추출하는 방법을 사용하였다. 이 연구는 한국어 시소러스가 없기 때문에 영어 시소러스를 사용하여 정서가를 추출한 것에 의미가 있다. 하지만 한국어로 된 단어와 영어로 된 단어가 뜻이 같아도 조금씩 다른 정서를 표현하는 경우가 있기 때문에 서로 다른 언어들을 사전을 통해 변환하여 정서를 추출하는 방법은 한계가 있다.

이처럼 인공지능 분야에서는 정서 단어를 추출하기 위해 단어가 표현하는 정서 상태를 임의로 설정하거나 단어별 정서 정보를 제공하는 영어 시소러스를 사

용하여 변환하는 방법을 사용하였다. 한편, 심리학 분야에서는 정서 단어를 유사성 평정을 한 후 다차원척도법(MDS: Multi-dimensional Scaling)을 사용하여 몇 가지 내적 하위 차원으로 구조화하는 정서의 차원 이론에 관한 연구들이 많다.

개별 정서는 고유한 특징을 보유하고 있는 상호 독립적인 존재라기보다는 몇 개의 차원들의 조합으로 구성된다는 것이 정서의 차원 이론이다(박인조와 민경환, 2005). 연구자들 마다 차원의 종류에 대해서는 이견이 많았으나 일반적으로 가장 큰 차원은 쾌-불쾌이고 두 번째 차원은 각성과 관련된 차원이라는 점에 대해서는 이견이 없었으며 3차원 이상에서는 연구자들마다 이견이 많았다(김영아 등, 1998; 김진관 등, 1998; 박인조와 민경환, 2005; 유상 등, 2006). 여기서 연구자들마다 가장 중요한 차원으로 쾌-불쾌 차원을 발견했지만, 두 번째 차원으로는 각성(arousal), 주의-거부(attention-rejection), 수면-긴장(sleep-tension) 등의 차원을 발견했다.

두 번째 중요한 차원으로 각성과 유사하지만 조금씩 다른 차원을 발견한 것에 대해 Russell과 Barrett(1999)은 각성, 에너지, 긴장(tension), 활동(activity) 등으로 분석되는 차원이 활성화(activation) 차원의 다른 이름일 뿐이고 이들 사이엔 매우 유사한 관계가 있다고 하였다. 또한 본 논문의 기초가 된 박인조와 민경환(2005)의 연구에서도 쾌-불쾌와 활성화 차원을 사용했으므로 본 논문에서도 두 번째로 중요한 정서의 차원을 활성화 차원으로 설정하였다.

3. 정서 단어 추출을 통한 정서 분석

본 논문에서는 형태소 분석기와 정서 단어 데이터 베이스를 통해 정서 단어를 추출하고 정서 모델과의 사상을 통해 정서를 분석한다.

3.1. 정서 단어 데이터베이스

한국어 정서 단어 목록을 작성한 연구(박인조와 민경환, 2005)에서는 정서의 차원 모델을 사용하여 정서 단어를 ‘원형성(prototypicality)’, ‘친숙성(familiarity)’, ‘쾌-불쾌’, ‘활성화 수준’의 네 차원으로 분석하였다. 이 연구에서는 연세대학교 언어정보개발연구원의 ‘현대 한국어 어휘 빈도’ 자료집(서상규, 1998)을 기초로

하여 최근 10년 동안에 출판된 신문, 잡지, 소설 및 수필, 취미 및 교양, 수기 및 전기, 국어 교과서, 희곡 및 시나리오 등을 대상으로 최종적으로 434개의 정서 단어 목록을 작성하였다.

‘원형성’ 차원은 단어들이 정서 단어로서 얼마나 적당한지를 평가한 것이고, ‘친숙성’ 차원은 단어들이 얼마나 친숙하게 느껴지는지를 평가한 것이다. ‘쾌-불쾌’ 차원은 쾌, 불쾌 정서 경험에 대해 단어들이 쾌, 불쾌 정도를 얼마나 잘 함축하고 있는지를 나타낸 것이고, ‘활성화’ 차원은 단어들의 활성화 경험에 대해 평가한 것이다. 각 평가는 7점 척도로 측정됐다.

Russell과 Barrett(1999)의 정서원형모형(Circumplex Model of Affect)에서는 감정의 ‘쾌-불쾌’, ‘활성화’ 차원만 다루고 있으므로 본 연구에서는 박인조와 민경환(2005)의 연구의 정서 단어 목록에서 ‘원형성’과 ‘친숙성’ 차원은 배제하였고 ‘쾌-불쾌’와 ‘활성화’ 차원만 사용하였다(표 1).

박인조와 민경환(2005)의 연구에서는 ‘불쾌’ 단어가 ‘쾌’ 단어보다 많고, ‘활성’ 단어가 ‘비활성’ 단어보다 많았는데, 이것은 진화심리학적으로 지지되는 결과라고 설명하였다. 또한 정서 단어의 분석에서 ‘쾌-불쾌’ 차원은 잘 나오나 ‘활성화’ 차원은 잘 나오지 않는다. 그 이유는 정서 단어에 있어서 ‘쾌’와 ‘불쾌’는 극단의 수준을 잘 반영하지만 ‘활성화’에서는 극단의 수준을 잘 반영 하지 못하고 중간 수준에 몰려 있기 때문이라고 설명한다(박인조와 민경환, 2005).

이 연구에서 사용한 ‘현대 한국어 어휘 빈도’ 자료집은 구어체를 반영하지 못한 데이터이기 때문에 실생활에 많이 쓰이는 구어체 정서 단어를 반영하지 못한 단점이 있다. 하지만 본 연구에서는 문학작품을 대상으로 정서 단어를 분석하므로 데이터로 사용하기 적합하다. 이 연구의 정서 단어 목록은 총 434개로, 이것으로부터 정서 단어 데이터베이스를 구축하였다.

표 1. 정서 단어의 쾌-불쾌, 활성화 목록 예

단어	쾌-불쾌	활성화
가쁜하다	4.51	3.80
가련하다	3.11	3.17
가소롭다	2.48	3.91
가없다	3.00	3.33
가증스럽다	1.88	5.08
.....

3.2. 텍스트에서 형태소 분석

텍스트의 단어와 정서 단어 데이터베이스의 매칭을 위해서 텍스트의 형태소를 분석하여 체인과 용언의 기본형을 추출하였다. 이를 정서 단어 데이터베이스와 매칭하여 텍스트의 정서 단어들을 추출하였다.

형태소 분석은 강승식(2002)의 형태소 분석 라이브러리(KLT 버전 2.10b)를 사용하였다. KLT는 한국어 형태소 분석, 자동색인(색인어 추출), 한글 맞춤법 검사/교정, 복합명사 분해, 한글 자동 띄어쓰기 등의 기능을 제공하며 32비트용 윈도우 및 리눅스에서 동작하며 분석 정확도가 높고 속도가 빠른 장점이 있다.

3.3. 정서 모델

정서는 개인적인 경험이나 상태 같은 추상적 개념이기 때문에 수리적으로 나타내기 힘들다. 그래서 우리는 ‘기쁨’, ‘슬픔’, ‘분노’와 같은 정서 단어를 통해 정서를 표현한다. 하지만 정서 단어 또한 추상적인 개념으로, 같은 단어가 표현하는 정서라도 다른 정서 상태를 나타내는 경우가 있다. 예를 들어, 독을 가진 뱀이 다가올 때 느껴지는 ‘공포’의 정서와 극장에서 팝콘을 먹으면서 호러영화를 볼 때 느껴지는 ‘공포’의 정서는 같다고 할 수 없다(Russell, 2003). 심지어 Pinker(1997)는 ‘공포’로 표기되는 정서는 하나가 아니라 여러 개라는 주장까지도 했다.

또한 정서 단어는 언어와 그 언어를 사용하는 문화에 따라 차이가 있어 표현에 한계가 있다(Russell, 1991). 예를 들어 영어로 ‘시원섭섭하다’를 번역하면 ‘bittersweet’이 된다. 하지만 ‘bittersweet’은 직역하면 ‘달콤쌉쌀하다’는 뜻으로 ‘시원섭섭하다’와는 정서적으로 차이가 있는 상태나 개념을 나타낸다. 뿐만 아니라 한국어의 ‘추억’이나 ‘추억’의 일본어인 ‘오모이데(おもいで)’의 경우 해당 언어를 모국어로 쓰는 한국인이나 일본인에게 명백히 정서적인 느낌을 준다. 하지만 ‘추억’이나 ‘오모이데(おもいで)’를 영어로 번역하면 ‘기억’으로도 해석될 수 있는 ‘memory’가 되어 한국인이나 일본인에게는 정서적인 느낌이 줄어들거나 없어진다.

이러한 정서와 정서 단어의 추상성 때문에 정서를 측정 가능한 단위로 쪼개어 객관적으로 설명하려는 시도가 이루어져 왔다. 기본 정서(Basic Emotion)는 이러한 시도의 일환으로, 정서를 대표하는 단어들로 정

서를 분류하고자 하는 방법이다. 예를 들어 Ekman(1992)은 정서를 행복, 슬픔, 분노, 공포, 혐오, 놀람으로 된 6개의 정서 단어로 분류하였다. 또한 Plutchik(1962)은 8개의 정서 단어로 분류하였으며, Izard(1977)는 10개의 정서 단어로 분류했다. 하지만 기본 정서가 되는 단어들은 연구에 사용된 언어에 종속적이기 때문에 보편적인 기본 정서를 나타내기에는 한계가 있다.

이 중 Ekman의 분류법은 전 인류에는 공통되는 정서적 얼굴 표정이 있음을 가정하고, 얼굴 표정을 통해 분류되는 정서가 6가지가 있음을 밝힌 것이다. 따라서 Ekman의 분류법은 6개의 정서 단어를 이용해 정서를 분류하지만, 본질적으로는 6개의 얼굴 표정으로 분류한 것이며 그에 해당하는 영어로 된 정서 단어를 매칭한 것이다.

따라서 Ekman의 기본 정서는 보편적인 정서 분류 단위로서 사용될 수 있다. 하지만 Ekman의 기본 정서는 얼굴 표정을 이용한 분류이기 때문에 얼굴 표정에 드러나지 않는 정서는 표현할 수 없고, 얼굴 표정에 드러나는 정서라도 기본 정서에 포함되지 않는 정서들은 표현할 수 없다. ‘외롭다’와 같은 경우 얼굴 표정에 잘 드러나진 않지만 정서적인 단어이고, ‘비웃다’와 같은 경우 얼굴 표정으로 나타나지만 기본 정서에는 포함되지 않는다.

한편, 기본 정서같이 정서의 단순 분류에 그치지 않고 정서 간의 관계를 나타내기 위해 정서 단어를 다차원 공간에 사상하여 표현하는 정서원형모형이 있다. 정서원형모형은 정서를 쾌 상태와 각성 상태로 표현하는 Wundt(1896)의 쾌-각성 이론(Pleasure-Arousal Theory)을 기반으로 하는 모형이다. 특정 언어에서 대표가 되는 정서 단어를 추출하는 것을 목표로 하는 기본 정서와 달리, 정서원형모형은 쾌 상태와 각성 상태를 기준으로 정서 간의 관계를 분류하고 표현하는 것을 목표로 한다. 따라서 기본 정서에 비해 언어 선택에 따른 한계가 적고 대부분의 정서를 표현할 수 있다.

정서원형모형은 Wundt의 이론을 바탕으로 정서를 쾌-각성의 2차원으로 분류하는 것에 주안점을 둔 초기 모형(Osgood, 1966; Russell & Mehrabian, 1974; Smith & Ellsworth, 1985; Traxel & Heide, 1961)으로 시작해 여러 가지 모형들이 연구됐다. 초기 모형은 정서의 크기를 나타낼 수 없었기 때문에 이를 개선하기 위해 쾌 차원으로는 정서를 분류하고 각성 차원으로는 정서의 크기를 구분하는 모형(Daly, Lancee, &

Polivy, 1983), 정서의 크기를 나타내는 축을 추가해 3차원으로 나타내는 모형(Clore, Ortony, & Foss, 1987; Shaver, Schwartz, Kirson, & O'Connor, 1987; Smith & Ellsworth, 1985), 축 간의 교차점과의 거리를 통해 크기를 나타내는 모형(Reisenzein & Schonpflug, 1992) 등 여러 가지 모형들이 제시되었다.

Russell의 정서원형모형은 대표적인 정서원형모형의 하나로서 정서를 쾌, 활성화의 2차원으로 분류한 초기 모형에서부터 축 간의 교차점과의 거리를 통해 크기를 나타내는 최근의 모형까지 꾸준히 연구되고 발전되어왔다.

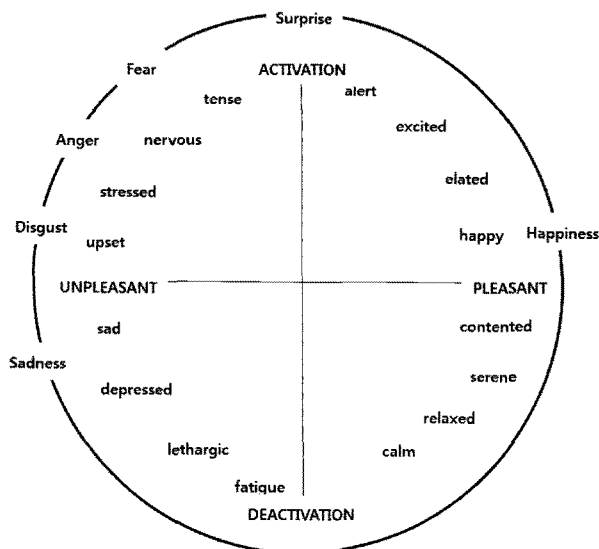


그림 1. 정서원형모형(Russell & Barrett, 1999)

그림 1은 Russell의 정서원형모형을 나타낸다. 가로축은 쾌(pleasant)-불쾌(unpleasant)의 정도를 나타내고 세로축은 활성화(activation)-비활성화(deactivation) 정도를 나타낸다. 특정 정서 상태는 이 두 축에서 각각 어떤 값을 갖는가에 따라 결정된다. 이 때, 해당 좌표가 어떤 정서를 나타내는가는 핵심 정서(Core Affect)를 이용해 나타낸다.

핵심 정서는 그림 1에서 원 내부 표기되어 있는 정서 단어들로, Russell이 정서원형모형 상에서 정서 상태를 언어적으로 분류하기 위한 기준이다. 원의 경계에 표기되어 있는 정서 단어들은 Ekman의 기본 정서가 정서원형모형 상에서 어디에 위치하는지를 나타낸다. Ekman의 기본 정서는 정서원형모형 상에서 제 2사분면(그림 1의 좌측 상단)에 집중적으로 분포하고 제 4사분면(그림 1의 우측 하단)에는 대응되는 정서가 없다. 대응되는 정서가 불쾌-활성화 부분에

집중되어 있는 것은 인류가 진화하는 과정에서 쾌 정서보다 불쾌 정서가 더 필요했고 정서가 표현되기 위해서는 비교적 높은 활성화 상태를 요구하기 때문이다(Johnson-Laird & Oatley, 1989; Shaver et al., 1987). 또한 본 논문의 정서 데이터베이스로 사용되는 연구(박인조와 민경환, 2005)에서도 불쾌-활성화 차원에 위치한 정서가 다른 차원에 위치한 정서보다 빈도가 높았다.

본 연구에서는 문학 작품의 정서 흐름을 다차원 공간상에 나타내기 위해 Russell의 정서원형모형을 사용한다. 또한 구체적인 정서의 변화보다는 지배적 정서를 추출하기 위해 16개로 구성된 Russell의 핵심 정서 대신 정서원형모형에 표현된 Ekman의 6가지 기본 정서를 이용하여 지배적 정서 흐름을 나타낸다.

3.4. 정서 단어 추출 방법

그림 2는 본 연구에서 지배적 정서를 분석하는 과정을 나타낸다. 먼저 문학 작품에서 텍스트를 입력하고 형태소 분석기를 이용하여 형태소를 분석하여 단어를 추출한다. 추출한 단어를 정서 단어 데이터베이스와 매칭하여 정서 단어를 다시 추출하고, 추출된 정서 단어의 쾌-불쾌와 활성화 값을 정서원형모형에 사상한다. 사상된 정서 단어와 정서원형모형 상의 기본 정서 단어를 매칭한다. 사상된 모든 정서 단어와 매칭된 모든 기본정서를 통해 지배적 정서를 분석한다. 이러한 과정을 윤오영의 수필 ‘방망이 깎던 노인’의 한 대목을 통해 살펴보면 그림 2와 같다.

먼저 그림 2의 (가) 문학 작품 입력에서 ‘방망이 깎던 노인’의 한 대목인 “나는 그 노인에 대해서 죄를 지은 것 같은 피로움을 느꼈다”를 입력한다. (나) 형태소 분석에서는 입력된 텍스트를 형태소 분석기를 통해 분석한다. 분석 결과는 ‘나’, ‘날다’, ‘노인’, ‘대하다’, ‘죄’, ‘짓다’, ‘지은’, ‘것’, ‘같다’, ‘피로움’, ‘피륙’, ‘느끼다’가 된다. (다) 정서 단어 데이터베이스를 통한 정서 단어 추출에서는 분석된 형태소를 정서 단어 데이터베이스와 매칭하여 추출한다. 그러면 분석된 형태소에서 ‘피륙’만이 매칭되어 추출된다. 정서 단어 데이터베이스 상에서 ‘피륙’에 대한 쾌-불쾌와 활성화 값은 각각 1.93과 3.85이다.

(라) 정서원형모형 사상에서는 쾌-불쾌와 활성화 값을 정서원형 상에 사상한다. (마) 기본정서 매칭에서는 사상된 정서 단어와 정서원형 상의 모든 기본 정

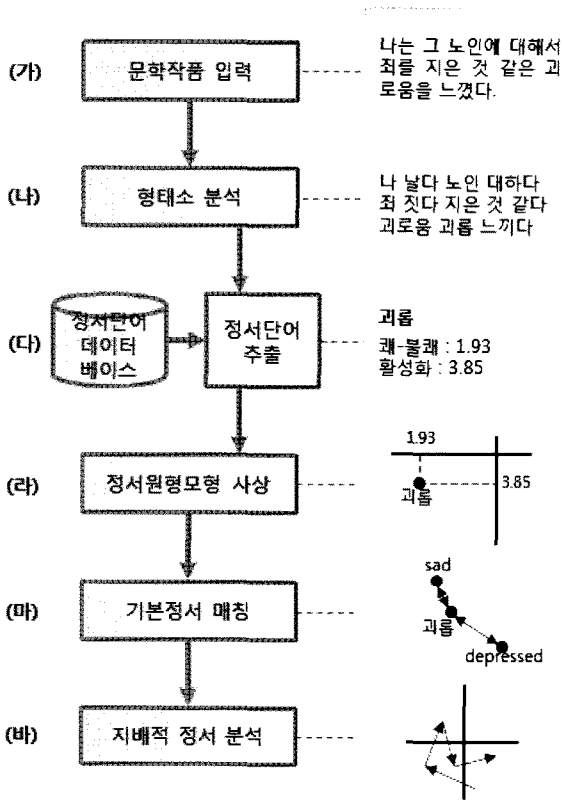


그림 2. 지배적 정서 분석 과정

서와의 거리를 비교한다. 이 때 사상된 정서 단어와 핵심 정서 간의 거리가 가장 짧은 것이 선택되는데 여기서는 'sad'가 선택된다. (바) 지배적 정서 분석에서는 모든 사상된 정서 단어와 모든 선택된 기본 정서를 통해 지배적 정서의 흐름을 분석한다.

다음 장에서는 이러한 추출 방법을 이용하여 두 문학 작품에서의 지배적 정서를 분석한다.

4. 지배적 정서 분석 및 평가

본 연구에서는 실제로 문학 작품의 지배적 정서를 분석하기 위해 현진건의 단편 소설 '운수 좋은 날'과 윤오영의 수필 '방망이 깎던 노인'을 분석 대상으로 선택 했다. 이 두 작품을 선택한 이유는 다음 세 가지이다.

먼저 두 작품 모두 중학교 교과서에 수록되어 친숙한 작품이므로 분석된 지배적 정서에 대한 검증이 용이하다. 둘째로, 길이가 짧아 문학 작품 전체를 분석하는데 무리가 없다. 셋째로, 내용이 짧음에도 불구하고 전개에 따른 정서적인 변화가 뚜렷하여 명확한 결과를 얻을 수 있다.

현진건의 단편 소설 '운수 좋은 날'의 줄거리는 다

음과 같다. 가난한 인력거꾼인 김침지는 아내가 중태인데도 불구하고 약값은 커녕 끼니조차 이어가지 못한다. 그러던 어느 날 운수 좋게도 많은 손님들이 모여들어 평생 보지도 못할 큰 돈을 얻는다. 그래서 오랜만에 술도 마시고 설렁탕을 사가지고 기분 좋게 먹을 것을 사가지고 집에 돌아갔으나 병든 아내가 죽어 있었다.

윤오영의 수필 '방망이 깎던 노인'의 줄거리는 다음과 같다. 화자가 방망이를 깎아 파는 노인에게 방망이를 깎아달라고 했는데 그 노인이 화자가 전차 시간을 놓치든 말든 아무리 재촉을 해도 자기 멋대로 느긋하게 깎아준다. 그래서 기분이 무척 나빴으나 집에 돌아와서 아내에게 방망이를 주니 아내가 무척 잘 깎은 방망이를 사왔다고 기뻐하여 화자는 노인에게 미안한 마음이 든다. 그래서 술이라도 대접하려고 노인을 다시 찾아갔으나 만나지 못한다.

표 2. 작품별 정서 단어 분석

	단어 개수	형태소 개수	정서 단어 개수
운수 좋은 날	7,622	5,280	18
방망이 깎던 노인	2,208	1,438	12

(단위 : 개)

표 2는 두 작품이 포함하고 있는 단어 개수와 형태소 분석 결과 얻어진 형태소 개수, 정서 단어 데이터베이스와 매칭 결과 얻어진 정서 단어 개수를 나타낸다. '운수 좋은 날'은 총 7,622개의 단어로 이루어져 있고 5,280개의 형태소로 분석되며, 정서 단어 데이터베이스와 매칭된 단어는 18개이다. '방망이 깎던 노인'은 총 2,208개의 단어로 이루어져 있고 1,438개의 형태소로 분석되며, 매칭된 단어는 12개이다. 매칭된 정서 단어에 대한 상세 내용은 표 3과 같다.

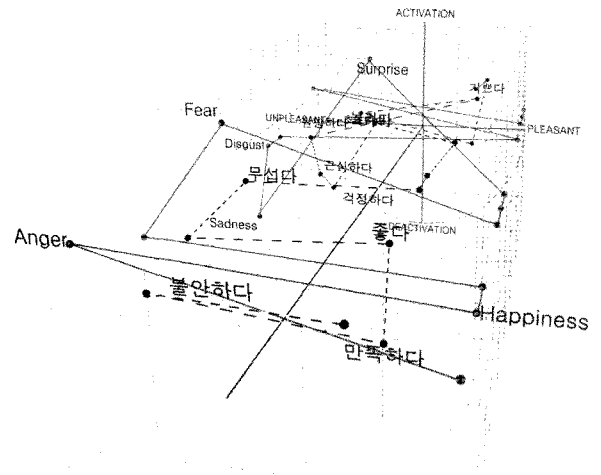
표 3에서 '좋다'의 경우 두 작품에서 모두 다수 매칭 되었는데 여기서 '좋다'는 정서 단어라고 말할 수 있는 기분이 '좋다'는 물론이고 운수가 '좋다', 방망이가 '좋다' 같이 정서 단어가 아닌 것까지 매칭되었다. 또한 정서 단어 데이터베이스에는 표정과 관련된 정서 단어가 없어서 '울다', '웃다'와 같은 정서 단어는 매칭되지 않았다.

그림 3은 매칭된 정서 단어를 정서원형모형에 사상한 것을 나타낸다. 점선으로 이루어진 경로는 표 4의 매칭된 정서 단어가 어떤 위상 경로를 가지는지를

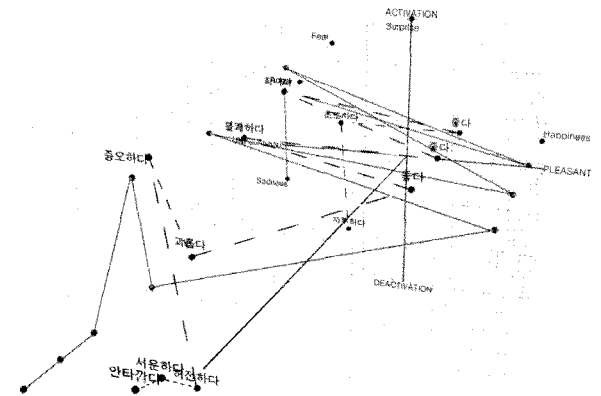
표 3. 매칭된 정서 단어

추출한 형태소	정서 단어 데이터베이스	쾌-불쾌 값	활성화 값
운수 좋은 날			
좋다	좋다	5.54	5.13
기쁘다	기쁘다	5.94	5.56
만족	만족하다	5.64	3.98
초조	초조하다	2.74	4.88
기쁘	기쁘다	5.94	5.56
원망	원망하다	1.86	4.63
근심	근심하다	3.24	4.99
걱정	걱정하다	2.94	3.94
놀라다	놀라다	4.12	5.67
기쁘다	기쁘다	5.94	5.56
좋다	좋다	5.54	5.13
좋다	좋다	5.54	5.13
무섭다	무섭다	2.58	5.39
원망	원망하다	1.86	4.63
좋다	좋다	5.54	5.13
만족	만족하다	5.64	3.98
불안	불안하다	2.51	4.88
좋다	좋다	5.54	5.13
방망이 깎던 노인			
지루하다	지루하다	2.64	2.14
초조	초조하다	2.74	4.88
좋다	좋다	5.54	5.13
화	화나다	2.17	5.91
좋다	좋다	5.54	5.13
불쾌	불쾌하다	2.13	5.46
좋다	좋다	5.54	5.13
괴롭	괴롭다	1.93	3.85
증오	증오하다	1.75	5.85
허전하다	허전하다	2.91	2.53
서운하다	서운하다	2.84	3.18
안타깝다	안타깝다	2.95	3.52

단어는 12개이므로 그림 3의 (가) ‘운수 좋은 날’의 경우 깊이축의 길이가 18이고 (나) ‘방망이 깎던 노인’의 경우 12이다.



(가) 운수 좋은 날



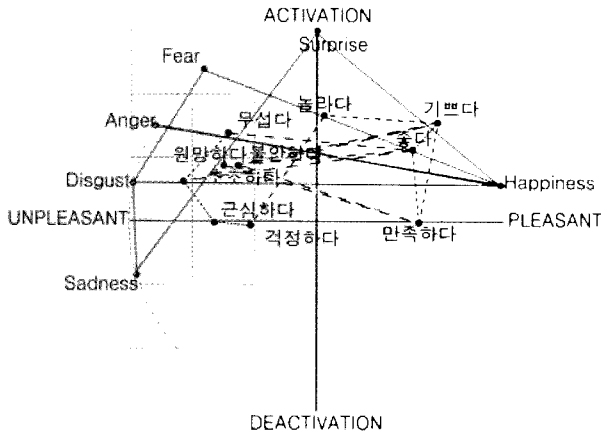
(나) 방망이 깎던 노인

그림 3. 정서원형모형 사상

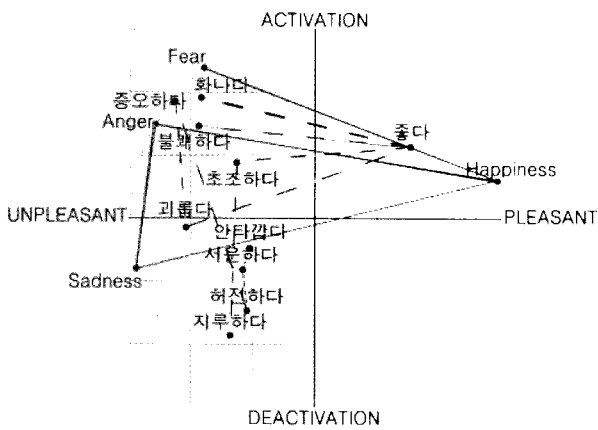
나타낸다. 또한 실선으로 이루어진 경로는 각 정서 단어와 가장 가까운 기본 정서를 나타낸다.

그림 3에서 가로 축과 세로 축은 정서원형모형 상의 쾌-불쾌와 활성화 상태를 나타내고 깊이 축은 추출된 정서 단어의 순서를 나타낸다. 정서 단어 데이터베이스 상의 쾌-불쾌 정도와 활성화 정도는 1에서 7까지의 값을 가지므로 쾌-불쾌, 활성화 축은 각각 1에서 7의 값을 가진다. 또한 ‘운수 좋은 날’에서 매칭된 정서 단어는 18개이고 ‘방망이 깎던 노인’에서 매칭된 정서

그림 4는 그림 3의 정서원형모형 사상을 깊이 축의 방향에서 바로 본 그림이다. 그림 1의 Russell의 정서원형모형에 대응하는 Ekman의 기본 정서를 보면 제 4사분면에 대응되는 정서가 없었다. 그런데 그림 4의 ‘운수 좋은 날’과 ‘방망이 깎던 노인’에서 추출된 정서 단어를 사상한 결과에서도 제4사분면(그림 4의 우측 하단)에는 대응되는 정서가 없다. 이러한 결과는 정서 단어 데이터베이스에 존재하는 단어들이 불쾌-활성화 차원에 높은 빈도를 가지고 있기 때문인 것으로 추측된다.



(가) 운수 좋은 날



(나) 방망이 깨던 노인

그림 4. 정서원형모형 사상 정면도

표 4는 사상된 정서 단어에 대응하는 기본 정서를 나타낸다. ‘좋다’, ‘기쁘다’, ‘초조하다’ 등은 ‘Happiness’에 매칭되었고 ‘초조하다’, ‘불안하다’, ‘불쾌하다’ 등은 ‘Anger’에 매칭되었으며 ‘원망하다’, ‘근심하다’는 ‘Disgust’에 매칭되었다. ‘걱정하다’, ‘지루하다’, ‘허전하다’ 등은 ‘Sadness’에 매칭되었고 ‘무섭다’, ‘화나다’는 ‘Fear’에 매칭되었으며 ‘놀라다’는 ‘Surprise’에 매칭되었다.

‘화나다’의 경우 의미상 ‘Anger’에 매칭되어야 할 것 같지만 ‘Fear’에 매칭되었다. 이는 ‘Fear’와 ‘Anger’ 사이의 거리가 상대적으로 가깝고 Russell의 모형에 Ekman의 모형을 대응시킬 때 사용된 표본 집단과 정서 단어 데이터베이스 구축에 사용된 표본 집단이 다르기 때문으로 추측된다.

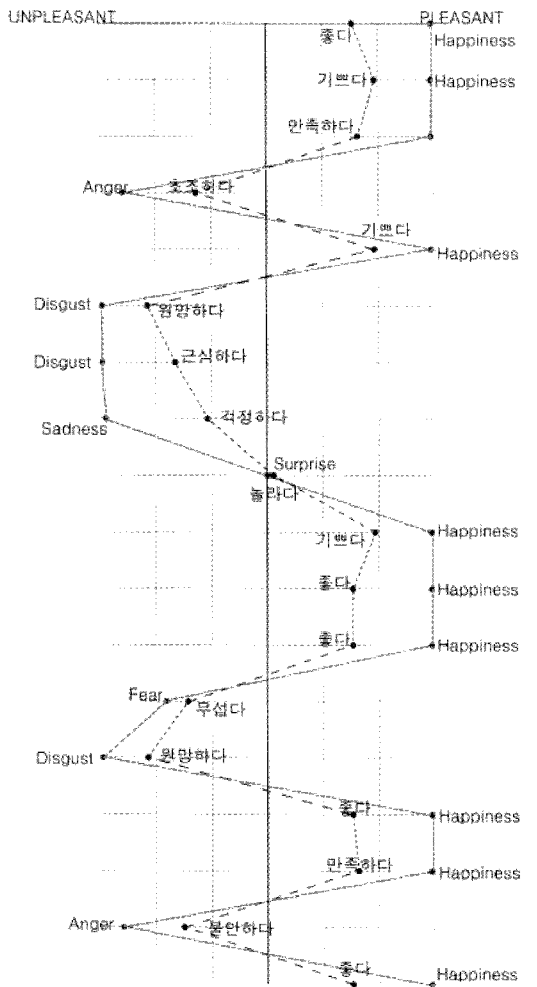
표 4. 정서 단어에 매칭된 기본 정서

순번	사상된 정서 단어	기본 정서	순번	사상된 정서 단어	기본 정서
운수 좋은 날					
1	좋다	Happiness	10	기쁘다	Happiness
2	기쁘다	Happiness	11	좋다	Happiness
3	만족하다	Happiness	12	좋다	Happiness
4	초조하다	Anger	13	무섭다	Fear
5	기쁘다	Happiness	14	원망하다	Disgust
6	원망하다	Disgust	15	좋다	Happiness
7	근심하다	Disgust	16	만족하다	Happiness
8	걱정하다	Sadness	17	불안하다	Anger
9	놀라다	Surprise	18	좋다	Happiness
방망이 깨던 노인					
1	지루하다	Sadness	7	좋다	Happiness
2	초조하다	Anger	8	괴롭다	Sadness
3	좋다	Happiness	9	중요하다	Anger
4	화나다	Fear	10	허전하다	Sadness
5	좋다	Happiness	11	서운하다	Sadness
6	불쾌하다	Anger	12	안타깝다	Sadness

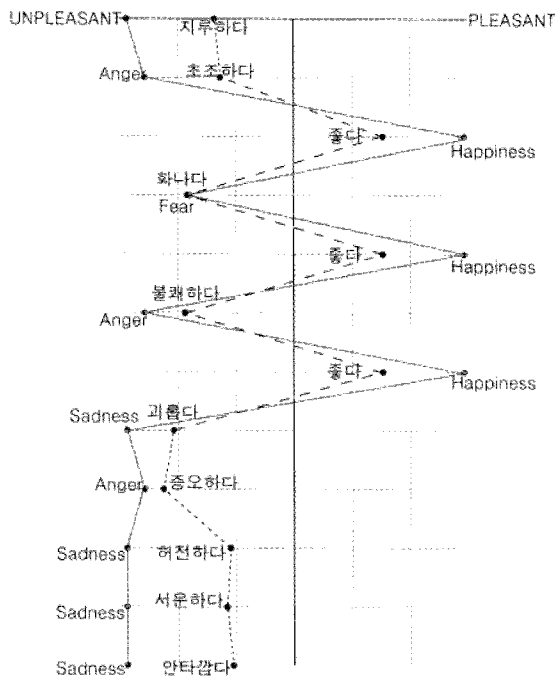
그림 5는 정서원형모형 사상을 세로 축의 방향에서 바라본 그림이다. 가로 축이 쾌-불쾌 축이고 세로 축은 매칭된 감정 단어의 순서이다. (가) ‘운수 좋은 날’은 전체 18개 기본 정서 중에 7개가 불쾌고 11개가 쾌로써 전체적으로 불쾌 정서보다 쾌 정서가 지배적이었다. 또한 쾌에서 불쾌로, 불쾌에서 쾌로 변환되는 영교차가 8번 발생했다. 이는 전체 17번의 변화 중 거의 반을 차지하는 횟수로 감정의 기복이 다소 심함을 의미한다.

실제로 ‘운수 좋은 날’에서는 생활고에 따른 불쾌 정서, 운수 좋게 일이 많음에 대한 쾌 정서, 아내의 병으로 인한 불안에 대한 불쾌 정서, 오랜만에 많은 돈을 번 것에 대한 쾌 정서 등 심한 감정 기복을 보여 준다.

그림 5의 (나) ‘방망이 깨던 노인’은 전체 12개 기본 정서 중 9개가 불쾌 정서이고 3개가 쾌 정서로 전체적으로 불쾌 정서가 지배적이다. 영교차는 6번 발생하였다. 하지만 3개의 ‘Happiness’를 매칭시킨 ‘좋다’는 정서적인 의미의 ‘좋다’가 아니라 ‘방망이가 좋다’, ‘사람 좋다’의 ‘좋다’로 원래는 영교차가 발생하지 않았어야 한다. 실제로 ‘방망이 깨던 노인’에서는 바쁜 데 느긋한 노인으로 인한 불쾌 정서, 오히려 화를 내는 노인에 대한 불쾌 정서, 노인에게 막 대했던 것에



(가) 운수 좋은 날



(나) 방망이 깎던 노인

그림 5. 정서원형 모형 사상 평면도

대해 미안한 불쾌 정서 등 불쾌 정서가 지배적이다.

그림 6은 정서원형모형 사상을 가로 축의 방향에서 바라본 이다. 가로 축은 매칭된 감정 단어의 순서이고 세로 축은 활성화 축이다. 그림 6의 (가) ‘운수 좋은 날’은 전체 18개 기본 정서 중 활성화가 17개, 비활성화가 1개로 활성화 정서가 지배적이다. 초반 7개 기본 정서는 활성화이고 8번째에 비활성화가 되었다가 그 이후로는 지속적으로 활성화가 된다.

실제로 ‘운수 좋은 날’은 전체적으로 활성화된 지배적 정서를 가진다. 초반부터 비 내리는 거리를 철썩거리며 뛰어다니고 아픈 아내를 때리는 등 활성화된 정서를 보여준다. 중간 즈음에 이르러 아내에 대한 걱정으로 비활성화 된 정서를 잠시 보여주다가 많은 돈을 벌고 술에 취해 흥청거리다 죽은 아내를 부여안고 목매여 우는 활성화된 정서를 보여준다.

그림 6의 (나) ‘방망이 깎던 노인’은 전체 12개 기본 정서 중에 활성화가 7개, 비활성화가 5개로 비활성화에 비해 활성화가 지배적이다. 또한 처음에는 비활성화였다가 이후 6개는 활성화가 되고 7개 째에 비활성화가 되었다가 8개 째에 활성화가 된다. 9개 째 부터는 비활성화를 유지한다. 이러한 흐름은 ‘방망이 깎던 노인’이 비활성화된 정서로 시작하였다가 활성화된 정서를 거쳐 비활성화된 정서로 끝남을 의미한다.

실제로 ‘방망이 깎던 노인’에서는 과거를 회상하는 담담한 이야기로 시작해 노인에 대한 불쾌함과 버스 시각으로 인한 초조함으로 인해 활성화된 정서를 보여준다. 그러다 방망이를 칭찬하는 아내를 보고 노인에 대한 미안함, 그리고 노인을 만나지 못한 것에 대한 안타까움으로 인한 비활성화된 정서로 마무리된다. 이러한 결과들을 정리하면 다음과 같다.

표 5는 ‘운수 좋은 날’과 ‘방망이 깎던 노인’의 지배적 정서 분석 결과를 보여준다. ‘운수 좋은 날’은 불쾌-활성화가 지배적이고 전체적으로 활성화된 정서를 가지며 쾌-불쾌의 기복이 심하다. 불쾌-활성화가 지배적이면서 활성화가 압도적이므로 격렬하면서 부정적인 정서를 가진다. 또한 동시에 쾌-불쾌의 기복이 심하여 쾌 또는 불쾌의 정서를 증폭시키는 구조를 갖는다. 그리고 전체적으로 활성화를 띠므로 작품 전반적으로 팽팽한 긴장감을 유지한다.

‘방망이 깎던 노인’은 ‘불쾌-활성화’가 지배적이다. 쾌-불쾌의 경우 영교차 6회로 기복이 심한 것으로 나왔지만 잘못 매칭된 ‘좋다’를 제외할 경우 영교차 0회의 지속적인 불쾌 정서를 갖는다. 또한 비활성화로 시

어 데이터베이스를 이용하여 정서 단어를 추출하였다. 또한 감정원형모형에 정서 단어를 사상하여 텍스트의 정서의 변화를 분석하였다.

제안한 분석 방법론을 현대 소설인 ‘운수 좋은 날’과 수필인 ‘방망이 깎던 노인’에 적용한 결과 쾌-불쾌, 활성화에 대한 지배적 정서 흐름과 변화 정도를 알 수 있었다. 다만 ‘방망이 깎던 노인’의 경우 정서적인 ‘좋다’와 정서적이지 않은 ‘좋다’가 구별되지 않아 지배적 정서의 쾌-불쾌 값과 세부 정서의 구체적인 변화 부분에 있어서 다소 잘못된 분석 결과를 보여줬다. 그럼에도 불구하고 지배적 정서와 지배적 정서의 흐름은 도출 가능했다.

본 논문의 결과가 가지는 두 가지 의미는 다음과 같다. 첫째, 심리학의 정서 연구 결과인 정서의 차원 모델을 사용하여 쾌-불쾌, 활성화 차원으로 정서 단어를 분석하고 이를 텍스트 분석 방법으로 알고리즘화했다. 둘째, 단위 시간 동안 주로 작용하는 정서인 지배적 정서 개념을 제안하여 문학 작품에 응용하여 분석하였다. 그 결과 문학 작품의 전반에 걸친 지배적 정서를 알 수 있었고 정서 변화와 흐름을 알 수 있었다.

본 논문에서 사용된 정서 단어 데이터베이스의 기초가 되는 박인조와 민경환(2005)의 연구에서는 ‘웃다’, ‘울다’와 같이 표정과 관련된 단어는 빠져 있다. 그럼에도 불구하고 표정과 관련된 단어는 정서적으로 민감하고 강한 정서성을 띄며, 문학작품에서 자주 사용된다. 따라서 정서 단어 데이터베이스에 표정과 관련된 정서 단어를 추가한다면 지배적 정서를 보다 명확하게 추출할 수 있을 것이다.

또한 본 논문에서는 두 편의 문학 작품만을 분석했는데, 더 많은 문학 작품들에 본 연구의 방법론을 적용한다면 문학 작품들을 유사한 지배적 정서 흐름별로 분류하는 것이 가능 할 것이다. 이러한 분류법은 제목, 장르, 작가 등의 기존의 검색 카테고리 외에 새로운 검색 카테고리 적용될 수 있을 것이며, 검색과 추천 시스템에 적용하여 보다 효과적인 시스템을 구현할 수 있을 것으로 기대된다.

본 논문의 방법론은 문학작품 외에 블로그, SNS, 신문기사 등의 텍스트로 확장시켜 지배적 정서를 분석하는데 사용될 수 있다. 이를 위해선 문어체에 한정된 데이터베이스를 구어체, 축약어, 통신체, 이모티콘과 같은 곳에서 사용되는 정서 단어를 포함하도록 확장해야 한다.

참고문헌

- 강승식 (2002). *한국어 형태소 분석과 정보검색*. 서울: 홍릉과학출판사.
- 김영아, 김진관, 박수경, 오경자, 정찬섭 (1998). 정서 관련 어휘 분석을 통한 내적 상태의 차원 연구. *감성과학*, 1(1), 145-152.
- 김진관, 문혜신, 오경자 (1998). 감성 개념 이차원 구조의 안정성. *감성과학*, 2(1), 43-52.
- 문현구, 장병탁 (2001). 채팅 텍스트로부터 화자 감정 상태 학습. *한국정보과학회 2001년도 봄 학술발표논문집*, 28(1-B), 340-342.
- 박인조, 민경환 (2005). 한국어 감성단어의 목록 작성과 차원 탐색. *한국심리학회지: 사회 및 성격*, 19(1), 109-129.
- 서상규 (1998). *현대 한국어의 어휘 빈도*, 서울: 연세대학교 언어정보개발연구원.
- 유상, 조정자, 한광희 (2006). 손동작과 정서 차원 분석. *감성과학*, 9(2), 119-132.
- 황재원, 고영중 (2008). 감정 자질을 이용한 한국어 문장 및 문서 감정 분류 시스템. *정보과학회논문지: 컴퓨팅의 실제 및 레터*, 14(3), 336-340.
- Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundation of the affective lexicon. *Journal of Personality and Social Psychology*, 53, 751-766.
- Daly, E. M., Lancee, W. J., & Polivy, J. (1983). A conical model for the taxonomy of emotional experience. *Journal of Personality and Social Psychology*, 45, 443-457.
- Ekman, P. (1992). An argument for basic emotion. *Cognition and Emotion*, 6, 169-200.
- Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.
- Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, 3, 81-123.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Osgood, C. E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology*, 7, 1(30).
- Plutchik, R. (1962). *The emotions : Facts, Theories and*

a New Model. New York: Random House.

- Plutchik, R. (1997). The circumplex as a general model of the structure of emotions and personality. In R. Plutchik & H. R. Conte(Eds.), *Circumplex models of personality and emotions*, Washington DC: American Psychological Association.
- Reisenzein, R., & Schonpflug, W. (1992). Stumpf's cognitive-evaluative theory of emotion. *American Psychologist*, 47, 34-45.
- Russell, J. A., & Mehrabian, A. (1974). Distinguishing anger and anxiety in terms of emotional response factors. *Journal of Consulting and Clinical Psychology*, 42, 79-83.
- Russell, J. A. (1991). Culture and the categorization of emotion. *Psychological Bulletin*, 110, 426-450.
- Russell, J. A., & Barrett, L. F. (1999). Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant. *Journal of Personality and Social Psychology*, 76(5), 805-819.
- Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, 110(1), 145-172.
- Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52, 1061-1086.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813-838.
- Traxel, W., & Heide, H. J. (1961). Dimensionen der Gefühle [Dimensions of emotions]. *Psychologische Forschung*, 26, 179-204.
- Wundt, W. (1896). *Grundrifi der Psychologie [Outlines of psychology]*. Leipzig, Germany: Engelmann.

원고접수 : 11.03.01

수정접수 : 11.04.15

게재확정 : 11.04.20