감정기반 정보 검색시스템에 관한 연구

A Study on Emotion based Information Retrieval System

김 명 관(Myung-Gwan Kim)*
박 영 택(Young-Taek Park)**

목 차

초 록

인터넷의 확산과 더불어 엄청난 사용자의 증가는 인터넷을 단순히 정보 검색의 대상으로만 삼는 것이 아니라 일반인들의 여가 문화를 즐기는 장이 되어가고 있다. 이와 같은 요구로 감정기반 문서 검색 및 분류 시스템을 제안한다. 이 시스템을 ECRAS라고 부른다. 감정 성분 추출은 로젯의 시소러스와 워드넷을 통해 이루어졌다. 감정 성분을 추출한 문서는 k-NN 기법을 기반으로 검색을 수행한다.

ABSTRACT

In this paper, we propose a document clustering and retrieval tool which allows users to manage their emotion based document access. This system name is ECRAS(Emotion based Clustering and Retrieval Agent System). Our system extract 5 emotion feature which like HAPPY, SAD, ANGRY, FEAR, DISGUST from various document. And, our system have retrieve documents for user query base on emotion feature.

* 서울보건대학 조교수
** 숭실대학교 교수
접수일자 1998년 11월 5일

1. 서론

1993년 미국 일리노이 대학에서 개발된 WEB 검색기 모자익의 출현 이후 인터넷 사용자는 폭발적으로 증가하고 있다. 미국의 경우 PC 통신이 모두 인터넷으로 통합되고 있으며 대표적인 통신업체인 AOL(America On Line)의 경우 약 1200만명의 가입자가 인터넷과 PC 통신을 활용하고 있다. 국내의 경우도 200만명이 넘는 사람들이 인터넷 전자우편과 월드 와이드 웹 검색을 사용하고 있다. 이와 같은 엄청난 사용자의 증가는 인터넷을 단순히 정보 검색의 대상으로만 삼는 것이 아니라 일반인들의 여가 문화를 즐기는 장이 되어가고 있다. 1996년 10월 10일에서 11월까지 한달 간 실시된 GVU(Georgia Instute of Graphics, Visualization and Usablility Center)의 15,000명을 상대로 얻은 조사 결과(이두희, 1997)에 따르면 인터넷 사용이유에 따른 분포는 다음과 같다. 복수 선택을 하였을 때 구경을 목적으로 사용하는 경우가 77.8%, 재미 63.79%, 교육이 53.29%, 일이 50.29%로 나타나고 있다.

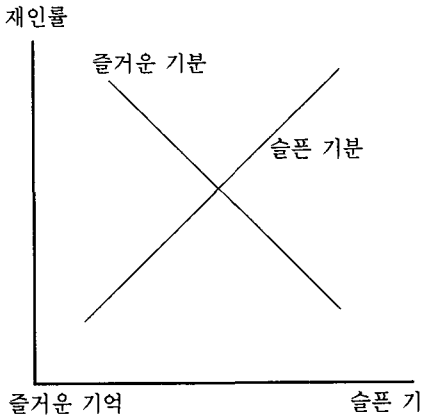
감정 표현은 위 항목 중 두 번째 위치를 차지하는 재미와 밀접한 관계를 갖는 언어 도구이다. 현재 우리말에 대한 감정 표현 시소러스에 대한 연구는 매우 미약한 실정이다. 다음 21세기에는 인터넷을 이용한 수많은 오락 서비스가 이루어질 것이며 이를 위해 감정표현을 지원하는 시소러스(Thesaurus)의 개발과 우리말 감정표현 검색기의 필요성이 대두될 것이다.

1983년 카네기 멜론 대학의 바워와 코헨은 '상태종속 기억장치'라는 이론을 제시하였다.(Clark, 1982) 그들에 의하면 개인의 감정은 정보 검색의 선택적 여과기(Selective Filter)와 기억 공간의 검색에 영향을 준다고 주장한다. 실험을 통해 그들은 〈그림 1〉과 같은 그림을 얻을 수 있었다.

이와 같은 실험 결과를 바탕으로 본 논문에서는 감정 기반의 정보 검색과 분류 기법들을 제안한다. 이렇게 구현된 결과로 다음과 같은 예의 정보 검색 및 분류가 가능하다.

1) "무서운 이야기가 있는 사이트는?"
2) "슬픈 영화를 찾아 주세요"
3) "1996년에 있었던 즐거운 사건들은?"

본 연구는 위와 같은 질의어를 사용할 수 있는 인터넷 정보 검색기를 구현하는 것을 목적으로 한다. 이를 구현하기 위하여 감정 성분을 해당 문서로부터 추출하여야 한다. 본 논문에서는 언어학자인 로젯(Roget,



〈그림 1〉 카네기 멜론 대학의 실험결과

1979)의 시소러스와 프린스턴 대학의 워드넷(Miller, 1993)을 사용하였다. 또한 추출된 감정 요소들을 가지고 사용자의 키워드와 감정 요구에 따른 검색을 수행한다. 이때 k-NN(k-Nearest Neighbor) 기법을 적용하여 적합한 문서를 찾는다. 실험을 위해 인터넷에서 구한 1,000여 개의 CNN 기사와 18편의 세익스피어 원작의 희곡, 100여 편의 최근 개봉했던 영화 대본을 가지고 실험을 수행하였다. 실험 결과 크기가 작은 문서로 이루어져 있는 웹 문서 등의 검색에 감정 요소 분류가 효용성이 있음을 보였다.

2장에서는 감정 요소를 사용하는 관련 연구들을 살펴보았다. 3장에서는 감정 성분 추출에 관한 내용을 다루었다. 본 논문에서 제안한 감정 기반 문서 분류 및 검색시스템을 ECRAS(Emotion based document Clustering and Retrieval Agent System)이라고 부르며 본 시스템의 구성에 대하여 다루었다. 4장에서는 추출된 감정값을 가지고 사용자의 질의에 대하여 검색을 실시하는 과정을 설명하였으며 5장에서는 영화 시나리오 등을 사용한 실험 결과를 보였다. 6장에서 결과와 향후 연구과제에 대하여 기술하였다.

2. 관련 연구들

정보처리에 있어서 감정에 대한 연구는 인공지능 분야의 중요한 과제였다. 대표적인 감정 처리를 포함한 시스템은 1981년 Colby의 PARRY(Colby, 1981), 1983년 Dyer의 BORIS(Dyer, 1983)와 1987년 OdEd(Dyer, 1987), 1991년 Reeves의 THUNDER (Reeves, 1991), 1987년에 발표된 ACRES (Reeves, 1987), 1992년 CMU의 Oz(CMU, 1992), 1997년 Wright(Wright, 1997)의 감정 행위자(Emotional Agents) 등이 있다.

PARRY는 '공포', '분노', '불신' 세 가지 감정 요소를 지원하는 시스템으로서 영어와 비슷한 입력형식을 가졌다. 주로 정신병 환자와의 대화를 목적으로 설계되었으며 악의적인 질문이 발견되면 물리적인 공포와 심리적인 분노로부터 불신에 관한 변수 값을 변경하여서 대응하는 결과 값을 출력하는 구조였다. BORIS는 유명한 인공지능 사례인 '이혼'에 관한 에피소드를 인식하는 심층인식(In depth understanding) 소프트웨어이다. 문맥 이해를 목적으로 설계되었으며 감정은 이야기 분석에 단서로서 사용된다. 감정을 위하여 6개의 스롯을 두어 변화를 처리하였다. 단순히 문맥 이해의 관점으로 감정요소를 다루었으므로 원칙, 목표, 성향 등의 구별을 갖지 않았다. 즉, 죄수가 탈옥 후 목표는 달성했지만 원칙과의 괴리 때문에 괴로워하는 감정의 변화를 표현할 수 없었다.

역시 Dyer의 OdEd는 신문의 사설을 이해하기 위한 시스템으로서 Shank의 CD (Conceptual Dependency) 이론을 적용하여 구현하였다. 지식베이스를 가지고 일종의 목표 기반 추론을 수행하였다. 즉, 실망을 만나면 지금의 목표는 취소하고 정지되어 있는 목표들 중에 다른 것을 선택하게 된다. 1991년에 나온 Reeves의 THUNDER는 역시 문서 구조를 이해하기 위한 시스템으로서 윤리 패턴을 가지고 이야기의 진행을 표현하고자

하였다. 예를 들어 장난으로 토끼를 죽이려던 사냥꾼이 도망가던 토끼에게 던진 다이너마이트 때문에 자신의 자동차가 폭파되는 이야기에서 윤리적 인과를 추론한다. 그러나 다양한 상황에 따른 윤리적 해석이 미흡하다는 평을 들었다. 예를 들어 낙선한 의원이 자살을 선택한다면 무사도 정신에서는 옹호되겠지만 천주교 교리로는 용서받지 못하는 다중 상황이 존재한다. Oz는 CMU네서 개발된 가상현실과 대화식 소설을 작성하기 위한 시스템이다. 이 시스템에서 생성된 가상 행위자들은 서로 경쟁적 행위를 하며 과정을 통해 환경 설정을 변화시킨다. 각각이 목표지향적인(Goal Directed Behavior) 행위를 하여 적절한 감정 변수를 갖는다. 이 시스템은 OCC(Ortony, Collins, Clore) 들의 모델을 기반으로 하고 있다. 각 감정 요소들이 발생하는 기저를 다음과 같이 표현하고 있다.

- JOY Goal Success
- DISTRESS Goal Failure
- HOPE Prospect of goal success
- FEAR Prospect of goal failure
- LOVE Attention to like object

그러나 Oz는 각 감정 요소의 정도를 표현할 수 없었으며 같은 타입의 감정들 사이의 차이를 보여주지 못한다. 위와 같이 살펴본 감정 요소를 사용한 기존의 시스템들은 주로 추론과 문장 이해를 목표로 구현된 것이다. 그러나 문장 이해는 대단위의 기반 지식을 갖추어야 얼마만큼의 결과를 보여줄 수 있다. 따라서 제한된 응용에서의 성과와는 달리 일반적인 범주에의 적용은 어려움을 갖는다. 본 논문에서 제안하는 ECRAS는 "전체 문장을 대표하는 감정값은 해당 감정 단어의 빈도에 비례한다"는 정의에서 출발한다. 정보 검색에서의 자동 색인 원리와 일치한다. (정영미, 1986) 따라서 본 시스템은 각 감정 요소들의 정도를 부여할 수 있으며 5가지(행복, 슬픔, 노여움, 공포, 혐오 : OCC 분류) 감정 요소들의 조합에 의해 해당 문서의 감정 대표 값을 산출하게 된다. 기존의 시스템들과 달리 본 논문에서 제안하는 감정 기반 분류 및 검색시스템은 일반적인 인터넷 환경의 웹 문서 및 뉴스, 소설, 영화 등의 사이트 분류 및 검색에 응용할 수가 있다.

3. 감정 성분 추출

본 연구의 첫 번 째 목표는 감정표현 단어들에 대한 시소러스를 구축하는 것이다. 정보검색 분야의 요구사항에 의해 언론연구원 등에서 한국어 시소러스 구축에 대한 많은 시도가 있었지만 전문분야 단어들에 대한 것이지 감정표현 단어들에 대한 시소러스 구축은 없었다. 감정 단어들에 대한 연구의 대표적인 결과물은 영국에서 나온 로젯(Roget)의 시소러스(Roget, 1979)와 미국 프린스턴 대학의 워드넷(Miller, 1993)을 들 수 있다. 로젯의 시소러스는 클레스 6에 감정(AFFECTION)을 일반적인 감정 단어, 개인적인, 동정, 도덕적인 것, 종교적인 감정 단어들로 분류하고 각각 단어군들의 유사어들을 280개의 군으로 나누어 정리해 놓고 있다. 1976년 밀

러와 존슨에 의해 만들어진 워드넷(WordNet)은 심리학과 언어공학이 만들어낸 결과물이다. 1993년 새로 구성된 결과물에 의하면 95,600개의 단어 구성으로 이루어져 있으며 51,500개의 단순 단어와 44,100개의 연어로 이루어져 있다. 70,100개의 단어 의미가 있으며 유사어가 있다. 이 워드넷은 5개의 범주로 구성되어있는데 명사, 동사, 부사, 형용사, 기능어로 되어있다. 예를 들어 명사인 경우 다음과 같이 구성되어 있다.

〈명사〉

57,000 단어 + 48,800 의미

25 개의 주제 화일 : 행위, 동물, 사람, ...

12 번 주제 :감정(398)

예)

04787051 12 n 01 sentimentality 0 002 @

04786928 n 0000 ~ 04787172 n 0000 | ...

sentiment mawkishness

위의 예는 워드넷의 각 단어가 상위어, 하위어, 유사어 등에 대한 연결을 네트웍으로 구성하고 있음을 나타낸다. 즉 04787051은 단어 sentimentality의 고유번호이며 @은 상위어가 04786928이며 ~은 하위어가 04787172라는 의미이다. 또한 맨 뒤에는 유사어들이 작성되어있다.

본 연구에서는 감정 단어들을 행복, 슬픔, 노여움, 공포, 혐오(Happy, Sad, Angry, Fear, Disgust) 등 5가지 범주로 나누고 각 범주의 단어들을 워드넷과 유사한 망 형태로 구성하였다. 또한 각 감정 성분들은 로젯 시소러스의 유사어 분류를 사용하였다. 구성한 결과 행복은 1,700 단어, 슬픔은 1,300 단어, 노여움은 1,400, 공포는 1,300 혐오는 1,500개로 구축했으며. 이와 같은 기본 단어 군과 워드넷의 관련 단어들을 연결하여 감정 성분 단어들의 시소러스를 구축하였다. 특히 위의 워드넷을 이용하여 로젯의 시소러스에서 가져온 분류 단어들에 유사어와 상위어 하위어 등을 본 시스템의 시소러스로 구성하였다. 감정 범주 분류기는 아래 그림 2와 같은 모습으로 구현된다. 구축된 시소러스 중 행복과 슬픔에 관련된 일부를 보면 다음과 같다.

〈HAPPY〉	〈SAD〉
- 858	-- 649
hope	badness
desire	hurtfulness
fervent	virulence
trust	bane
confidence	plague spot
reliance	insalubrity
faith	hoodoo
belief 484	Jonah
affiance	malignity
assurance	malevolence
secureness	tender mercies
reassurance	ironically
promise	ill-treatment
anticipation	annoyance
expectation 507	molestation
hopefulness	abuse
buoyancy	oppression
optimism	persecution
enthusiasm	outrage
	misusage
	injury
	damage
	knockout drops
	badness
	peccancy
	abomination
	painfulness
	pestilence
	disease

위의 내용 중 단어 옆에 있는 숫자는 상위어 및 하위어를 명시한다. 따라서 시소러스에서의 위치에 따라 누적 값이 차등으로 적용된다. 기본적으로는 로젯의 시소러스를 사용하였으며 워드넷의 상위어 하위어 부분을 일부 포함 시켰다. 또한 감정성분 추출 방법은 다름과 같다.

1) 입력된 문서에서 하나의 단어를 추출한다.(Poter의 알고리즘을 사용하여 복수형, 진행형 등을 표준형으로 바꾼다.)
2) 불용어인 경우 이를 제거하고(불용어 사전으로) 아니면 시소러스를 검색한다.
2) 시소러스를 사용하여 해당 감정성분 값을 누적한다.
3) 값을 계산할 때 시소러스의 상위어 하위어 유사어 등에 값을 차등화 하여 누적한다.
4) 만들어진 5개의 벡터 값을 정규화 한다.

각 문서로 부터 위 시소러스를 사용하여 5개 성분의 벡터를 추출한다.[4] 문서 j의 추출된 성분 벡터는 다음과 같이 표현한다.

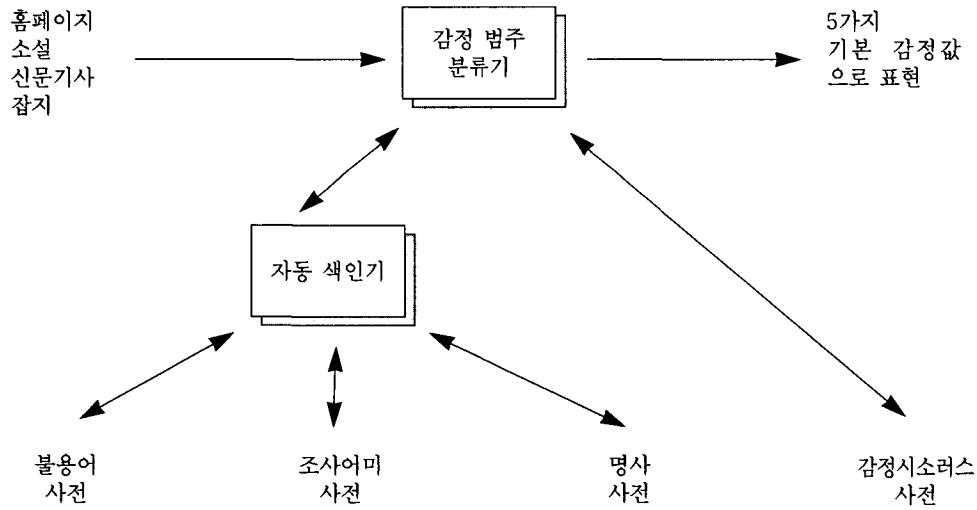
$$EVj = \langle V1j, V2j, V3j, V4j, V5j \rangle$$

$$Vij = \log(Fij) / \log(Fwj)$$

or

$$Vij = Fij / Fwj$$

위에서 Vij는 문서 j에서의 i번≳ 감정 성분을 나타내며 Fij는 i번째 감정 단어가 j문서에서 나타난 빈도이며 Fwj는 j문서에 있는 모든 단어의 개수가 된다. Vij는 0에서 1 사



〈그림 2〉 ECRAS의 감정 범주 분류기

이의 값을 갖는다.

4. k-NN 기반의 검색

k-NN(k-Nearest Neighbor)은 instant 기반 및 후 수행(Lazy) 학습의 대표적인 알고리즘이다.(Mitchell, 1997) 이 방법은 학습할 대상 개체들을 n 차원 공간 속의 한 점으로 인식하고 벡터화 된 좌표를 학습 시 단순하게 기억장치에 저장만 한다. 클래스를 모르는 질의어가 올 때 이를 n 차원 상의 벡터로 표현하고 점 사이의 거리 구하는 방식으로 가장 가까운 k개의 점들을 구하여 이들 점들이 가지고 있는 클래스 정보로 질의어의 클래스를 결정하는 기법이다. 이 경우 질의어 개체가 m개의 성분을 가진다면 다음과 같이 표현된다.

$\langle a1(x), a2(x), a3(x),.....,am(x)\rangle$

두 점들 사이의 거리는 다음과 같다.

$$d(xi,xj) = \sqrt{\sum_{r=1}^{m}(a_r(xi)-a_r(xj))^2}$$

위와 같은 거리 구하는 방식으로 질의어 xq에 대한 분류 값을 학습된 개체 중 가까운 거리의 x1, x2, ..., x_k 를 대상으로 다음과 같이 구한다.

$$f(x_q) = \underset{v \in V}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(v,f(x_i))$$

이 k-NN 기법은 단순하면서도 동작 과정을 쉽게 이해할 수 있고, 학습 시간이 빠르며 (ID3의 100배 정도) 높은 정확성(accuracy)으로 널리 사용되고 있다.

감정표현을 지원하는 인터넷 검색기는 기존의 인터넷 검색기(야후, 알타비스타 등에서 사용하는 일반 검색 엔진)와는 다르게 키워드 입력과 감정표현 입력기 두개의 입력을 가지고 검색을 실시한다. 감정표현 검색기의 작동 과정은 다음과 같다.

1) 5가지 기본 감정값을 사용자가 조절하여 원하는 감정 벡터를 만들게 된다. (행복: 0.1, 슬픔:0.8, 노여움:0.3, 공포:0.0, 혐오:0.0)
2) 사용자가 찾고자하는 기본 키워드를 입력한다.("영화")
3) 감정표현 검색기는 키워드를 가지고 일반적인 인터넷 역화일 검색을 실시한다.
4) 결과로 나온 문서들을 감정표현 시소러스를 사용하여 감정 벡터들을 구한다.(타이타닉 기사 => 행복: 0.2, 슬픔:0.6, 노여움:0.3, 공포:0.2, 혐오:0.4)
5) 5가지 기본 감정 벡터를 사용한 적합도 검사를 실시한다.(k-NN 기법을 사용)
6) 적합도가 가장 높은 문헌들의 리스트를 사용자에게 보여준다.

다음 예는 위에 보인 사용자 감정 요구와 타이타닉의 적합도 계산 과정을 보여준다.

적합도(t,q) = sum(t*q)/root(sum(t**2) * sum(q**2))
= (0.1*0.2+0.8*0.6+0.3*0.3)/ root((0.01+0.64+0.09)*(0.04 +0.36+0.09+0.04+0.16))
= 0.8309 ==〉 타이타닉은 사용자의 요구와 관련 있는 영화임!

또는 질의어와 문서 감정 요소 값 사이의 거리를 구한다.

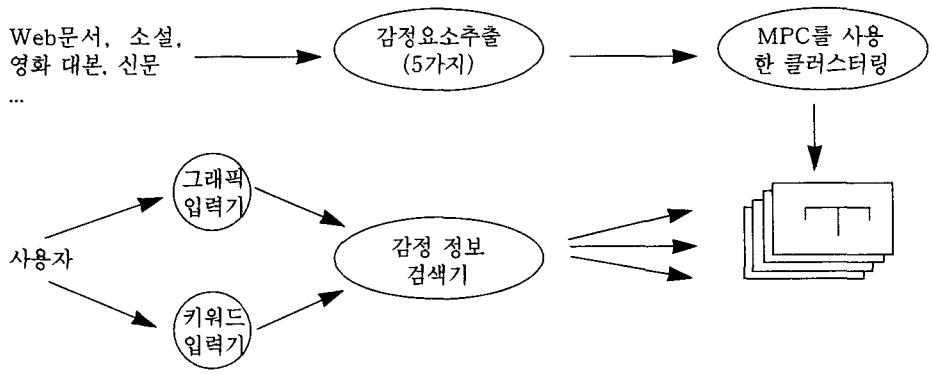
거리(t,q) = root((0.1-0.2)**2+(0.8-0.6)**2+(0.3-0.3)**2+(0.2)**2+(0.4)**2)
= 0.5 ==〉 다른 문서들과의 거리를 다 구한 후에 가장 적은 거리의 문서가 채택됨!

5. 실험 및 고찰

본 논문에서 제안한 ECRAS의 실제 응용을 위해 약 100여 편의 영화 시나리오와 세익스피어 작품 18개 CNN의 인터넷 기사 1000개를 사용하여 실험 환경을 구축하였다. 영화 시나리오는 주로 최근(1~2년)에 개봉한 작품들을 위주로 하였다. 이 문서들은 ECRAS에 입력으로 들어가며 결과는 감정 요소 추출기를 이용한 정보검색이 수행된다. 그 과정을 다음 〈그림 3〉으로 나타내고 있다.

ECRAS에서 감정 요소를 추출하는 과정은 〈사례 1〉를 통해 보여줄 수 있다. 키워드 자동 색인은 기존 Salton(Salton, 1989) 등의 일반적인 방법을 사용하였다.

구축된 실험 환경을 가지고 각각 감정 요소들을 추출한 후에 이들 요소들에 대한 문서 크기에 따른 표준편차의 차이를 구해 보았다. 표준편차가 클수록 사용자의 감정 요소에 따른 검색에 애매성이 적어질 것이다. 실험 결과 5,000 단어 미만의 짧고 간략한 신문 기사(1,000개) 등은 변별력 즉 표준편차의



〈그림 3〉 ECRAS

〈사례 1〉 ECRAS에서 감정 요소를 추출하는 과정

1) 감정 성분 추출

	전체 단어 수	행복*희망	슬픔	분노*공포
CNN의 Sexy스타 관련 기사	1465	32	11	28
CNN의 Clinton 관련 기사	2423	23	18	71
CNN의 "Simon Birth" 기사	1104	29	6	16
세익스피어 멕베드	18765	373	317	664
세익스피어 Tempest	17862	508	326	476
영화 Mr. Bean의 대본	30011	864	521	463
영화 Avenger의 대본	23837	603	317	560
영화 Sneakers의 대본	18983	483	184	281
영화 Brazil의 대본	34895	921	461	693

2) 정규화

CNN의 Sexy스타 관련 기사		0.021	0.007	0.019
CNN의 Clinton 관련 기사		0.009	0.007	0.029
CNN의 "Simon Birth" 기사		0.026	0.005	0.014
세익스피어 멕베드		0.019	0.017	0.035
세익스피어 Tempest		0.028	0.018	0.02
영화 Mr. Bean의 대본		0.028	0.017	0.015
영화 Avenger의 대본		0.025	0.013	0.023
영화 Sneakers의 대본		0.025	0.009	0.015
영화 Brazil의 대본		0.026	0.013	0.019

*위 실험에서 분노와 공포는 하나로 계산하였으며 시소러스 구축의 문제 때문에 감정요소 혐오는 배제하였다.

크기가 크게 나타났으며(표준편차 평균: 10.12) 상대적으로 큰 10,000 단어 이상의 영화 대본(100개)이나 세익스피어의 희곡 등(18개)은 감정 요소들 간의 차이가 크지 않게 나타났다.(표준편차 평균: 6.20) 따라서 감정 요소에 의한 문서 분류 및 검색은 인터넷의 웹 등 온라인 문서의 주류를 이루는 게시판 및 신문 기사나 잡지의 기고 내용 등 가볍게 읽을 수 있는 문서들이 적합한 것으로 보인다.

6. 결론 및 향후 연구 방향

본 논문에서는 감정 성분을 해당 문서로부터 추출하여 감정 성분을 기반으로 한 정보 분류 및 검색 시스템(ECRAS)을 제안하였다. 언어학자인 로젯(Roget, 1979)의 시소러스와 프린스턴 대학의 워드넷(Miller, 1993)을 사용하였으며 감정 성분 추출에 관한 내용을 다루었다.

관련 연구들에서 살펴본 것 같이 기존의 감정 처리에서 주로 다루었던 문장 이해는 대단위의 기반 지식을 갖추어야 얼마만큼의 결과를 보여줄 수 있다. 따라서 제한된 응용에서의 성과와는 달리 일반적인 범주에의 적용은 어려움을 갖는다. 본 논문에서 제안하는 ECRAS는 "전체 문장을 대표하는 감정값은 해당 감정 단어의 빈도에 비례한다"는 정의에서 출발한다. 정보 검색에서의 자동 색

인 원리와 일치한다. (정영미, 1986) 따라서 본 시스템은 각 감정 요소들의 정도를 부여할 수 있으며 5가지(행복, 슬픔, 노여움, 공포, 혐오 : OCC 분류) 감정 요소들의 조합에 의해 해당 문서의 감정 대표 값을 산출하게 된다. 기존의 시스템들과 달리 본 논문에서 제안하는 감정 기반 분류 및 검색시스템은 일반적인 인터넷 환경의 웹 문서 및 뉴스, 소설, 영화 등의 사이트 분류 및 검색에 응용할 수가 있다. 5가지 감정 성분의 추출로 나온 결과는 정도 값으로 대상 문서의 대표 감정값을 알 수 있다. 사용자는 이를 키워드와 자신이 찾고자 하는 감정 요소 값을 질의어로 정보검색을 할 수 있다.

실험을 위해 CNN의 기사들과 세익스피어 원적의 희곡들, 각종 영화 대본 등을 사용하였다. 실험 결과 각 감정요소들의 표준편차가 대상 문서의 크기에 반비례한다는 것을 보였다. 따라서 우리의 감정 분류 및 검색 시스템은 문서의 크기가 작은 인터넷 홈페이지 등의 검색 분류에 알맞을 것으로 보인다. 앞으로 우리말에 대한 시소러스 구축과 각 감정 성분들 간의 관계성 및 이를 사용자에게 쉽게 보여주고 사용하게 할 수 있는 방법론이 요구되어 진다.

참 고 문 헌

Aha, D. W. 1991. " Instance-Based Learning Methods", Machine Learning, 6, 37-66

Alpaydim, E. 1995. " Voting Over Condensed Nearest Neighbors" , Bogazici Univ

Alpaydim, E. 1996. "GAL : Networks that Grow When They Learn and Shrink When They Forget" , International Computer Science Institute, Berkeley: CA, TR-91-032

Clark, M S. 1982. "Affect and Cognition", LEA Publishers

Colby, M. 1981. "Modeling a paranoid mind", The Behavioral and Brain Sciences, 4(4) : 515-560

Dyer, M. G. 1983. "In depth understanding", MIT Press

Dyer, M. G. 1987. "Emotions and their computations" , Cognition and Emotion, 1(3) : 323-347

Elliot, C. D. 1992. "A Process model of emotions in a multi-agent system", Ph.D thesis, north-west Univ.

Fisher, D. 1987. "Knowledge Acquisition via Incremental Conceptual Clustering", Machine Learning, 2, 139-172

Hart, P. E. 1968, " The Condensed Nearest Neighbor Rule" , IEEE Transaction on Information Theory, 14, 515-516

Miller, G. A. 1993. "WordNet: An On-line Lexical Data Base", Hillsdale

Mitchell, T. 1997. "Machine Learning", McGraw-Hill

Reeves, J. F. 1991. " Computational morality : A process model of belief conflict and resolution for story understanding" , Technical Report UCLA-AI-91-05, UCLA AI Lab

Roget, P. M.. 1979. "Roget's Thesaurus", Gramercy Books

Salton, G. 1989. " Automatic Text Processing", Addison Wesley

Sestito, S. 1994. "Automated Knowledge Acquisition", Prentice Hall

Wright, I. P. 1997. "Emotional Agents", Ph.D thesis, Univ. of Birmingham

이두희. 1997. "인터넷 마케팅", 영진 출판사 : 264-273

유상진 외 3인. 1997. "현대 통계학", 법영사

정영미. 1986. "정보 검색론", 정음사 : 181-208