



국민대학교
전자정보통신대학
컴퓨터공학부

캡스톤 디자인 I

종합설계 프로젝트

프로젝트 명	심리상담 데이터를 이용한 텍스트 내 감정 분석 시스템
팀 명	트로스트
문서 제목	계획서

Version	1.2
Date	2018-03-08

이름	김동현, 김윤창
----	----------


문서 정보 / 수정 내역

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 "심리상담 데이터를 이용한 텍스트 내 감정 분석 시스템" 를 수행하는 팀 "트로스트"의 팀원들의 자산입니다. 국민대학교 컴퓨터공학부 및 팀 "트로스트"의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

수정날짜	대표수정 자	Revision	추가/수정 항 목	내 용
2018-03-08	김동현, 김윤창	1.0	최초 작성	계획서 초안 작성

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

목 차

1	개요.....	4
1.1	프로젝트 개요.....	4
1.2	추진 배경 및 필요성.....	4
2	개발 목표 및 내용.....	5
2.1	목표.....	5
2.2	연구/개발 내용.....	5
2.3	개발 결과.....	6
2.3.1	결과물 목록 및 상세 사양..... 오류! 책갈피가 정의되지 않았습니다.	
2.3.2	시스템 기능 및 구조..... 오류! 책갈피가 정의되지 않았습니다.	
2.4	기대효과 및 활용방안.....	7
3	배경 기술.....	7
3.1	기술적 요구사항.....	7
3.2	현실적 제한 요소 및 그 해결 방안.....	8
3.2.1	하드웨어..... 오류! 책갈피가 정의되지 않았습니다.	
3.2.2	소프트웨어..... 오류! 책갈피가 정의되지 않았습니다.	
3.2.3	기타..... 오류! 책갈피가 정의되지 않았습니다.	
4	프로젝트 팀 구성 및 역할 분담..... 오류! 책갈피가 정의되지 않았습니다.	
5	프로젝트 비용..... 오류! 책갈피가 정의되지 않았습니다.	
6	개발 일정 및 자원 관리.....	9
6.1	개발 일정.....	9
6.2	일정별 주요 산출물.....	10
6.3	인력자원 투입계획..... 오류! 책갈피가 정의되지 않았습니다.	
6.4	비 인적자원 투입계획..... 오류! 책갈피가 정의되지 않았습니다.	
7	참고 문헌..... 오류! 책갈피가 정의되지 않았습니다.	

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

1 개요

1.1 프로젝트 개요

데이터를 기반으로 감정 데이터베이스를 구축하고, 이를 모델링하여 감정 리포트를 개발한다.

감정 데이터베이스는 8가지의 대표 감정과, 그에 속하는 하위 표현, 그리고 개별 표현별 감정의 강도로 구성된다. 분류기는 텍스트 데이터 속에서 감정 어휘들을 추출하고, 문법 규칙을 적용하여 최종적으로 감정 종류와 강도를 도출한다.

모바일을 통해 사람들이 고민을 작성하면 텍스트에서 감정을 추출해 리포트 형태로 간단한 분석과 함께 제공한다.

1.2 추진 배경 및 필요성

대한민국 국민 4명 중 1명은 전 생애에 걸쳐 한 번 이상 우울, 불안 등 정신건강 문제를 경험한다. 만약 가벼운 우울감까지 포함한다면 그 수는 더 커질 것이다. 이에 따라 정신건강관리에 대한 사회적 관심과, 지불 의사는 계속해서 높아지고 있다. 특히, 1인 가구 폭증에 따라 ‘외로움 관리’(loneliness management)라는 이름으로 각종 소셜 미디어나 인공지능(AI)를 활용한 산업 역시 시장이 확대되고 있다. 개인이나 가정 뿐 아니라 직장 등 사회적 단위에서도 이를 효과적으로 관리하려는 움직임이 늘어나며, 심리 및 감정을 효과적으로 다루는 멘탈헬스케어 서비스의 수요가 급증하고 있다. 우리나라는 다른 나라에 비해 상대적으로 스트레스 지수가 매우 높은 편이나 이들 중 정신과 전문의, 기타 정신건강 전문가를 통한 상담 및 치료를 받은 비율은 15.3%에 해당하며, 이는 선진국의 정신보건 서비스 이용률에 비해 매우 낮은 수준이다. 실제로 정신과 병원이나 심리상담 센터에 가기에는 대면상담에 대한 부담, 높은 비용, 시공간의 제약 등의 부담으로 현실적으로 이용하기란 어렵다.

초기 치료시기를 놓쳐 정신질환이 만성화되거나, 쇼핑이나 폭식, 알코올로 스트레스를 풀다가 상황이 악화될 경우 극단적인 선택으로까지 이어질 수 있다. 10년 째 OECD 국가 중 자살률 1위를 지키는데다, 우울증으로 인한 사회경제적 비용이 2011년 10조 3826억 원에 달하는 데, 그를 사전에 예방하기 위한 솔루션은 미비한 상태다.

쉽게 접근할수 있는 스마트폰 어플리케이션을 도구로 정신과 진료, 심리치료 이전에 스스로가 감정을 정리하고 진단할 수 있는 솔루션을 통해 예방책을 마련하고 정신건강 치료에 대한 진입장벽을 낮추는데 그 목적이 있다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

2 개발 목표 및 내용

2.1 목표

- 대표감정 8개(기쁨, 사랑, 분노, 좌절, 슬픔, 공포, 수치, 연민)와 그 하위 표현들로 구성된 감정 사전 제작
- 감정별 2개의 강도 축을 도출하고, 이를 명명함.
- 텍스트 데이터 속에서 감정 어휘를 추출하여, 대표 감정 및 해당하는 강도 축을 도출하는 감정 분류기 개발

2.2 연구/개발 내용

- (1) 감정사전 제작
- (2) 감정 강도 지정
- (3) 사례 분석 유의점 도출

감정 분석 모델링


- (1) 상담 텍스트 전처리
- (2) 규칙 기반 감정 분류 모델
- (3) 감정 사전 데이터베이스 기반 감정 분류기

데이터 기반 감정 분류 모델

- (1) 텍스트 전처리를 위한 불용어 제거 및 어휘 사전 구축
- (2) 워드 임베딩
- (3) 감정 분류기 학습

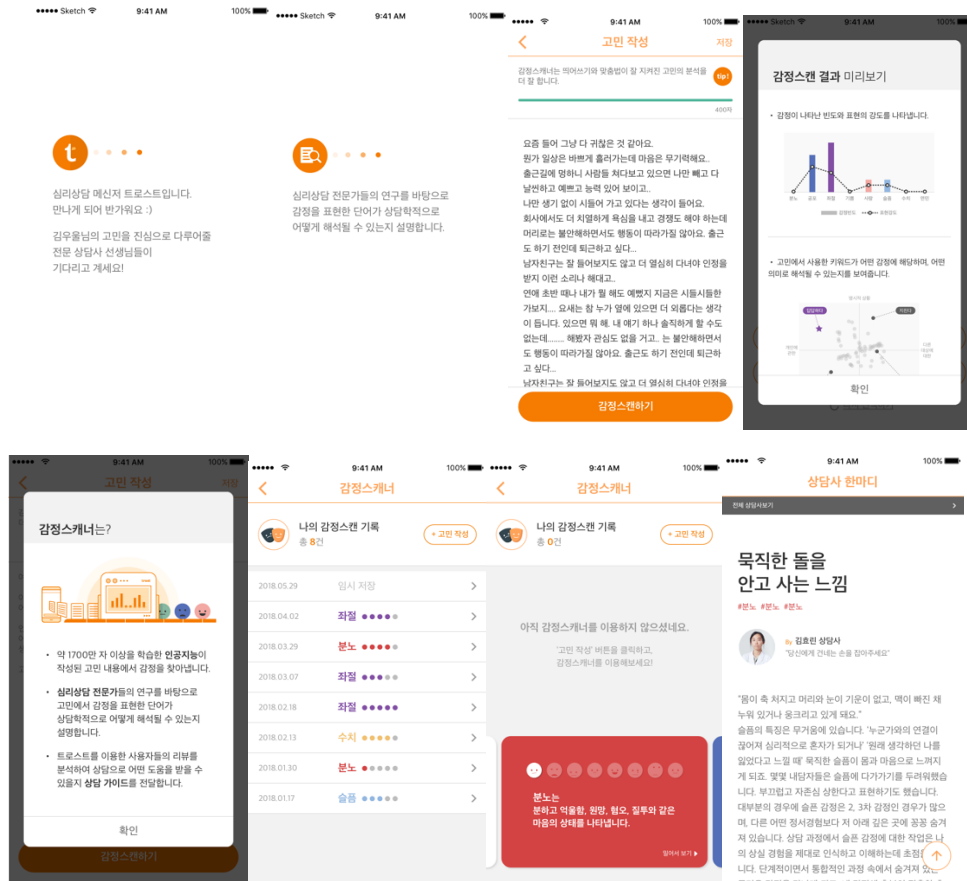
컴퓨터 시스템을 통해 텍스트를 기계적으로 분석하는 ‘텍스트 마이닝(Text Mining)’기술을 이용하여 사적 텍스트를 형태소별로 나눈 후, 그 내용(주제, 주제어, 동사), 문법(조사, 어미, 동사의 형태), 형식, 말투(어미의 변형), 이모티콘, 특수기호 등을 분석한다. 그리고 ‘내용 분석(Content Analysis)’기술을 이용하여 이 분석 결과가 다양한 감정들과 연계된 정도를 통계적으로 도출하여 최종적인 감정상태를 표시한다. 이를 통해 분석할 수 있는 감정은 ‘우울함’, ‘행복함’, ‘슬픔’, ‘화남’과 같은 개인적인 감정뿐만 아니라 관계에 있어서의 감정도 포함된다. 감정에 대한 분석결과는 그래프와 수치로 표현될 수 있다.

。

 <div> <p>국민대학교</p> <p>컴퓨터공학부</p> <p>캡스톤 디자인 I</p> </div>	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

2.3 개발 결과

감정 분석 시스템 예시



① 분류 정확성

개발기술(감정 분류기)이 도출한 감정 분석 결과와 전문 심리상담사가 사전에 작성한 감정 분석 정답 데이터 사이의 유사도를 측정함

② 도출 감정 수

- 8개의 대표감정과 그 하위감정으로 감정 데이터베이스를 구성함.

긍정	부정						
기쁨	사랑	분노	좌절	슬픔	공포	수치	연민

③ 내담자 및 상담사 만족도

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

컴퓨터 시스템을 통해 텍스트를 기계적으로 분석하는 ‘텍스트 마이닝(Text Mining)’기술을 이용하여 사적 텍스트를 형태소별로 나눈 후, 그 내용(주제, 주제어, 동사), 문법(조사, 어미, 동사의 형태), 형식, 말투(어미의 변형), 이모티콘, 특수기호 등을 분석한다.

그리고 ‘내용 분석(Content Analysis)’기술을 이용하여 이 분석 결과가 다양한 감정들과 연계된 정도를 통계적으로 도출하여 최종적인 감정상태를 표시한다.

이를 통해 분석할 수 있는 감정은 ‘우울함’, ‘행복함’, ‘슬픔’, ‘화남’과 같은 개인적인 감정뿐만 아니라 관계에 있어서의 감정도 포함된다. 감정에 대한 분석결과는 그래프와 수치로 표현될 수 있다.

2.4 기대효과 및 활용방안

기술 개발 시 온라인 심리상담과 같이 몸짓, 눈빛 등의 비언어가 전달되기 어려운 대화 속에서 감정을 효과적으로 포착할 수 있다. 또한 심리상담을 진행하며 변화되는 내담자의 감정 변화 추이를 확인할 수 있어 상담의 효과를 간접적으로 측정할 수 있다. 자동화된 솔루션은 개인이 언제 어디서나 본인의 고민이나 생각에 대한 감정을 심층적으로 확인할 수 있어 정신건강의 위기를 효과적으로 예방할 수 있게 하며, 데이터를 기반으로 상담사, 상담 기법, 상담 프로그램 등을 추천하는 내담자 맞춤형 서비스 제공이 가능하다. 기존 지인이나 친구 사이에서 이루어지던 대화에서 등장하는 감정의 깊이와 범위를 넘어서 데이터를 활용하여, 멘탈헬스케어 분야를 넘어서 기타 인공지능 컴퓨터 시스템에 데이터 솔루션으로서도 기능할 수 있다.

3 배경 기술


3.1 기술적 요구사항

형태소 분석

파이썬 한국어 오픈 소스 형태소 분석기 - konlpy

감정 사전 데이터베이스 기반 감정 분류기 - 꼬꼬마 형태소 분석기(kkma)

word2vec 감정 분류기와 데이터 기반 감정 분류 모델 - 트위터 형태소 분석기(twitter)

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

감정 분류기

파이썬 자연어 처리 오픈 소스- NLTK


머신러닝 패키지 - scikit-learn

워드 임베딩 - NLTK에서 제공하는 서포트벡터머신, 나이브 베이즈 분류 알고리즘

scikit-learn의 로지스틱 회귀, 앙상블 알고리즘(상당 텍스트 학습)

3.2 현실적 제한 요소 및 그 해결 방안

현재 정답 데이터에서 감정별 데이터 갯수가 적고 비율이 다르다 .따라서 낮은 정확도는 현재 정답 데이터에서 클래스의 불균형이 원인으로 생각되며 중복을 제외한 전체 데이터의 개수도 부족하기 때문에 학습이 정상적으로 이루어 지는데 제한이 있다고 생각한다. 태깅 데이터 갯수를 늘려서 학습량을 늘리면서 정확도를 높이고 룰베이스에서 데이터베이스로의 비중을 높이면서 해결 가능하다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명		
	팀 명		
	Confidential Restricted	Version 1.2	20xx-MAR-25

4 개발 일정 및 자원 관리

4.1 개발 일정

주차별	활동계획	예상 소요시간	기타
1주차	다차원 척도 분석	30시간	감정 강도 지정
2주차	코딩 자료 생성	30시간	
3주차	분석 결과 정리	30시간	
4주차	사례 분석 유의점 도출	30시간	
5주차	상당 텍스트 전처리	30시간	감정 분석 모델링
6주차	규칙 기반 감정 분류 모델	30시간	
7주차	문장 내 감정 주제 후보 추출	30시간	감정 사전 데이터베이스 기반 감정 분류기
8주차	주어 판단 확률 모델링	30시간	
9주차	문장 내 감정 주제 후보 추출	30시간	word2vec 기반 감정 분류기
10주차	주어 판단 확률 모델링	30시간	
11주차	각 분류기의 결과 융합	30시간	가중치 비교
12주차	텍스트 전처리를 위한 불용어 제거 및 어휘 사전 구축	30시간	데이터 기반 감정 분류 모델
13주차	워드 임베딩	30시간	
14주차	감정 분류기 학습	30시간	
15주차	A/B 테스트	30시간	마무리

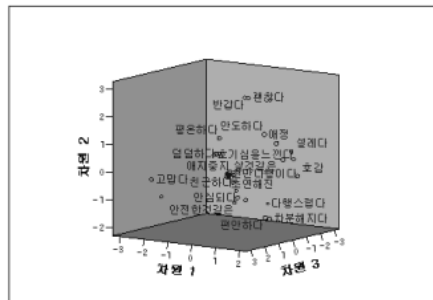
4.2 일정별 주요 산출물

대표감정 구분	이전 단어 수(개)	제외된 단어 예시	최종 단어 수(개)
좌절	238개	‘불길하다’, ‘심려하다’ 등	100
분노	207개	‘우습다’, ‘가슴에서 피가 끓다’ 등	100
공포	96개	‘해괴하다’, ‘기상천외하게 느껴지다’ 등	35
수치	94개	‘참회하다’, ‘회오’ 등	33
슬픔	93개	‘고적하다’, ‘비감에 젖다’ 등	48
사랑	50개	‘낯다’, ‘버틸만하다’ 등	35
기쁨	48개	‘흥나다’, ‘우와’ 등	30
연민	21개	‘가련하다’, ‘노파심’ 등	11

Results			
For matrix			
Stress =	.15910	RSQ =	.90341
Configuration derived in 2 dimensions			
Stimulus Coordinates			
	Dimension	1	2
Stimulus Number	Name		
1	동정하다	1.4068	-.4553
2	딱하다	1.3934	-.4523
3	불쌍하다	-2.5967	.4744
4	안타깝다	.3966	-1.9712
5	애듯하다	-.5305	-.0914
6	연민	.5820	1.4305
7	안쓰럽다	1.2421	1.2361
8	째하다	-.4521	-.1101
9	애처롭다	-.4806	-.0202
10	평하다	-.4806	-.0202
11	측은하다	-.4806	-.0202

유도된 자극의 위치

유클리디안 거리 모형



유도된 자극의 위치

유클리디안 거리 모형

