

기업 자기소개서 대상 텍스트 마이닝 적용 연구

김경식*, 김성보**, 김응모*
*성균관대학교 컴퓨터공학과
**성균관대학교 전자전기컴퓨터공학과
e-mail : {y2k3252, ksb31205, ukim}@skku.edu

A Study on the Application of Text Mining for Corporate Application form

Kyoung-Sik Kim*, Seong-Bo Kim**, Ung-mo Kim*
* Dept. of Computer Science, Sung-kyun-kwan University
** Dept. of Electronic Electric and Computer engineering, Sung-kyun-kwan University

요 약

최근 우리 나라 청년실업률의 증가와 함께 기업에서는 좋은 인재를 뽑기 위하여 정형화된 자격 증보다는 지원자의 경험과 역량을 보기 위한 자기소개서에 대한 중요성이 커지고 있다. 그리하여 비정형 데이터를 분석하는 대표적인 방법인 텍스트 마이닝 기법을 이용하여 취업 커뮤니티에 올라온 합격한 삼성, 현대자동차, LG 자기소개서 데이터를 얻어내고 그 후 KONLPY 패키지를 통하여 형태소 분석을 실시한다. 합격자소서에 자주 나온 단어의 순위를 매기고 공통적으로 많이 들어간 단어와 각 대기업 별 차이가 나는 단어를 회사의 인재상과 비교해본다. 그리고 취업 준비 생들에게 효율적인 방법을 통해 자기소개서를 작성하여 합격률을 높이는 방향으로 사용한다.

1. 서론

국내 청년실업률이 12%대를 넘어서 역대 최고를 기록하였다. 실업률도 6년 만에 가장 높은 수준에 달하면서 고용 관련 지표가 전체적으로 악화하는 모습을 보였다. 그리고 경기 침체 국면은 장기화 조짐을 보이고 있는 가운데 앞으로 3년간 취업시장에 쏟아질 대학 졸업자 인원이 역대 가장 많다. 2017년 졸업자수는 지난해 33만 명을 크게 웃도는 35만 명 이상으로 예상되고 있다. 그리고 기업들은 불황의 늪 속에 채용 규모를 줄이고 있다. 고용노동부 자료에 따르면 300인 이상 규모의 대기업은 올해 상반기 29,792명만 채용하겠다고 밝혔다. 최근 8년간 가장 낮은 수치로 같은 기간에 비해 8.8% 감소하였다[1].

그런 가운데 대기업에서는 서류전형으로 지원자들의 정형화된 사진, 어학성적, 해외경험 기입란을 삭제하고 지원자들의 경험과 인성, 역량을 보기 위한 자기소개서를 이용한 평가를 주로 하였다. 그러나 여러 그룹에서는 서로 다른 인재상을 말하고 있고 지원자들은 과연 그 인재상과 서류 전형에 통과된 기준이 같은지 궁금해한다.

본 논문에서는 이러한 점을 관찰하기 위하여 텍스트 마이닝을 이용하여 각 대기업 그룹사의 합격 지원서에서 어떤 단어가 많이 사용되었는지를 살펴보고 과연 그들의 인재상과 맞는지 비교 할 것이다.

2. 관련연구

2.1 텍스트 마이닝(Text Mining)

사람들은 기업 내에서 생산되는 데이터를 효율적이고 관리하기 쉽도록 하기 위해 데이터베이스에 정형화된 데이터를 저장해 왔고, 이렇게 저장된 데이터는 특정한 패턴이나 경향을 데이터마이닝 기술로써 재생산 활용되어 왔다. 하지만, 최근 기업에서 활용되는 데이터의 비중이 데이터베이스에 저장된 정형데이터 보다는 특정한 형식을 취하고 있지 않은 비정형데이터에 치중되고 있는 상황이다[2].

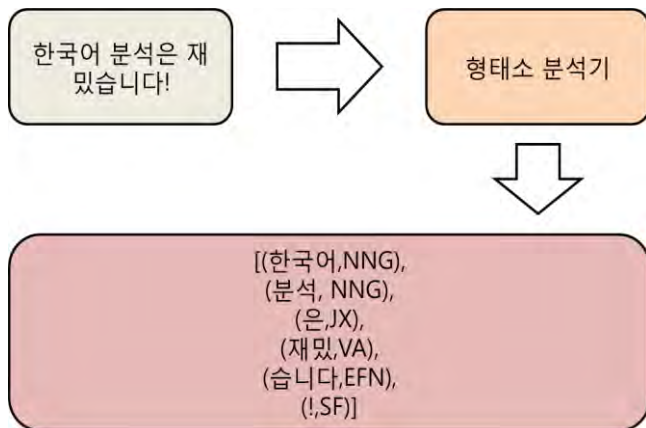
텍스트 마이닝은 이러한 비정형 데이터에 대하여 자연언어처리 기술과 문서 처리 기술을 적용하여 유용한 정보를 추출, 가공하는 것을 목적으로 하는 기술이다. 문서 요약, 문서 분류, 문서 군집, 특성 추출 등이 텍스트 마이닝의 핵심 연구 분야이며 그 응용 분야는 매우 다양하다.

데이터 마이닝 관점에서 문서로부터 구조화된 정보를 추출하여 데이터베이스화 시키거나 규칙을 찾아내는 것은 가장 일반적인 응용이며, 사용자가 Web 상에서 문서를 찾는 것을 도와주거나 문서에 쓰인 자연언어 식별, 문서 분류 정보를 이용한 문서 재해석 등이 가장 대표적인 응용 분야라 할 수 있다[3].

2.2 형태소 분석기

형태 분석이란 단어를 형태소 단위로 분리해내는 과정을 말한다. 다양한 품사와 의미를 갖는 자연어를 처리함에 있어 형태 분석은 가장 우선시되는 작업이라고 할 수 있다. 이러한 작업을 컴퓨터를 통해 자동화하

여 수행하는 소프트웨어를 형태소 분석기라 한다. 형태소 분석을 통해 나온 결과는 (그림 1)과 같다[4].



(그림 1) 'KoNLPY' 형태소 분석기의 입출력 예

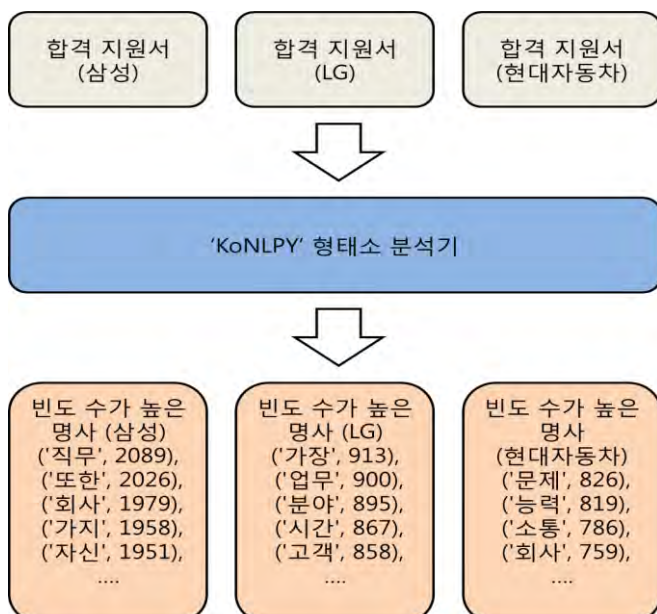
2.2.1 KoNLPY

KoNLPY(Korean Natural Language Processing in Python)는 Python의 범 언어적 자연어처리 패키지 NLTK와 마찬가지로, 자연어 처리를 갖 배우기 시작한 학생이나 자연어 처리를 연구 목적으로 사용하려는 연구자에게 적합하다. 이 도구를 통해 형태소 분석, 말뭉치 분석, 언어 처리 등이 가능하다. KoNLPY는 서울대학교 산업공학과에서 개발한 형태소 분석기의 Wrapper 패키지로서 이미 개발된 한글 형태소 분석기 세트를 Python 언어에서 사용가능 하도록 하였다[5].

3. 기업 자기소개서 대상 텍스트 마이닝 연구

3.1 기업 자기소개서 대상 텍스트 마이닝 연구

기업 자기소개서 대상 텍스트 마이닝 연구는 (그림 2)와 같은 방식으로 분석하였다. 그리고 합격자 지원서는 취업커뮤니티인 사람인에 올라와 있는 삼성, LG, 현대자동차 그룹을 통계 Tool인 R을 이용하여 수집하였다.



(그림 2) 자기소개서 텍스트 마이닝 분석 과정

텍스트 마이닝 과정에서 지원서에서는 명사만을 추출한다. 명사와 함께 동사 등을 추출하려 하였지만 형태소 분석기가 완전히 한국어의 동사를 구분하는 것이 불가능하다고 판단하였다. 불필요한 단어들이 순위에 올라 유의미할 결과가 나오지 않을 수도 있기 때문이다. 수집된 지원서들을 분석 후 결과가 나올 때 핵심 단어를 추출하는 과정이 필요하다. 형태소 분석기 특성상 '가장', '또한' 등 조사와 부사를 명사로 취급하는 경우가 있어서 그러한 단어들은 제외하였다.

3.2 연구 내용

삼성, LG, 현대자동차의 지원서를 분석한 결과 공통적인 부분도 있고 각 회사별로 차이점도 발견하였다. 공통적으로 가장 많이 나온 단어는 경험, 기술, 일, 결과, 사람, 목표, 팀, 업무, 시간, 고객, 직무, 도전, 과정, 생각이었다. 이를 바탕으로 같이 팀으로 일하는 사람과 직무에 대한 이해 신입사원으로써 보여줄 수 있는 도전과 생각 등을 공통적으로 추구하고 있다는 것을 알 수 있다.

3.2.1 LG

LG 그룹의 인재상을 보면 알 수 있듯이 세계 최고에 도전하는 사람, 창의적으로 일하는 사람, 끊임없이 혁신하는 사람, 정정당당하게 경쟁하는 사람의 인재상을 원하고 있다[6]. 또한 공통적으로 자주 나온 단어를 제외하고는 본인, 분야, 프로젝트, 역량, 공부, 설계, 열정, 지식, 제품, 공정, 문제 등의 단어가 자주 등장하였다. 자주 나온 단어들을 관찰하면 LG 그룹에서는 역량, 지식, 프로젝트 등 자신의 능력에 대한 설명이 합격 지원서에 많이 포함된 것을 알 수 있다.

3.2.2 현대자동차그룹

현대자동차그룹의 인재상을 보면 도전적 실행, 소통과 협력, 고객최우선, 인재존중, 글로벌 지향하는 인재상을 원하고 있다[7]. 그리고 지원서를 살펴보면 자동차, 사람, 생산, 자신 전공, 문제, 능력, 소통, 지식 등의 단어가 자주 등장하였다. 현대자동차 그룹의 인재상과 비교하면 소통과 사람에 대한 것을 중요시 여기는 것을 알 수 있다. 또한 자동차에 대한 관심이 중요하다는 것을 보여준다.

3.2.3 삼성그룹

삼성그룹의 인재상을 보면 인재제일, 최고지향, 변화선도, 정도경영, 상생추구를 하는 인재상을 원하고 있다[8]. 그리고 지원서를 살펴보면 사람, 성장, 노력, 팀, 도전 등의 단어가 자주 등장하였다. 삼성 그룹의 인재상과 비교하면 인재제일, 변화선도, 최고지향등과 함께 노력이라는 단어가 눈에 띈다. 앞에 나온 인재상들과 함께 노력하는 인재에 높은 점수를 준다는 것을 보여준다.

4. 결론 및 향후 연구

텍스트 마이닝을 이용하여 대기업이 공통적으로 원하는 인재상에 대해서 분석하고, 또한 각 대기업 그룹별로 유의미한 차이의 인재상을 보였다. 본 연구는 단순히 회사의 인재상과 문맥을 다듬어서 서류전형의 합격률을 높이는 수준만이 아닌 직접 합격 지원서를 분석하여 실질적으로 회사에서 원하는 인재가 무엇인지 알 수 있었다. 향후에는 이번에 제외하였던 품사들인 동사와 서술어까지 추가하여 더욱더 자세한 연구를 할 계획이다. 또한 합격되지 못한 지원서들을 데이터 셋으로 추가하여 합격한 지원서와 합격되지 못한 지원서들의 차이를 비교하면 기업 지원자들에게 어떠한 역량을 갖추어야 하고 어떠한 방식으로 지원서를 써야 하는지에 대해서 도움을 줄 수 있을 것이다.

참고문헌

- [1] 허인혜, “취업 빙하기...男 10~12 학번, 女 12~14 학번.”, 한스경제, 2017
- [2] 장선일, “텍스트마이닝을 활용한 고객 이탈모형”, 中央大學校 大學院: 統計學科 統計學專攻, 2011
- [3] 안태성, 이경일, 서형국. “텍스트마이닝 기반 고정밀 검색시스템”, 한국정보처리학회. 2004, p.89
- [4] 강승식, “한국어 형태소 분석과 정보검색”. 홍릉과학출판사, 2002
- [5] 박은정, “KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지”. 한글 및 한국어 정보처리 학술대회, 언어공학연구회, 2014
- [6] LG, ”Meet LG” 인재상: <http://meet.lg.com/learn/human/HumanResource.rmi>, 2017
- [7] 현대자동차그룹, “현대자동차그룹”. 인재상: <http://www.hyundai.co.kr/Careers/Talent.hub>, 2017
- [8] 캠퍼스 잡앤조이 온라인팀, “삼성전자 자기소개서는 어떻게 써야 할까? 삼성전자 자소서 문항 분석”. 캠퍼스 잡앤조이, 2016