

기업 리뷰 웹 사이트 텍스트 분석을 통한 직원 불만 표현 추출과 불만 원인 도출 및 해소 방안

백혜연¹ · 박용석^{2*}

Employee's Discontent Text Analysis on Anonymous Company Review Web and Suggestions for Discontent Resolve

HyeYeon Baek¹ · Yongsuk Park^{2*}

¹Graduate Student, Graduate School of Information Security, Sejong Cyber University, Seoul, 05000 Korea

^{2*}Professor, Graduate School of Information Security, Sejong Cyber University, Seoul, 05000 Korea

요 약

전현직 직원에 의한 산업정보 유출 비율이 80%에 이르나 산업정보유출 사고에 대한 뉴스기사나 정보유출 행위의 원인에 대한 연구들에서는 그 원인들을 처우나 인사 불만 등으로 간략하게 설명하고 있다. 본 연구에서는 전현직 직원들이 익명 기업리뷰 웹사이트에 남긴 기업에 대한 평가 텍스트를 분석하여 기업에 대한 불만 내용들을 더욱 구체적으로 확인하였다. 이 중 어떠한 불만사항이 퇴직이나 퇴사, 나아가 산업인력유출의 결과로 이어질 수 있는지 파악하기 위해 불만 분야에 대한 의미사전목록을 제시하고 부분문법그래프(LGG)를 구축하였다. 또한 텍스트 분석 결과에서 나타난 전현직 직원들의 불만사항과 기존 연구들에서 설문을 통해 정리한 인력유출 원인을 서로 비교하였다. 추가적으로 분석된 불만을 바탕으로 기업불만 해소를 통한 인력유출 방지 방안을 간략 제시하였다. 기존 설문 위주의 산업 인력 유출에 대한 분석에 더하여, 웹 크롤링을 통한 자유롭고 솔직한 불만 분석을 제공하는 데 의의가 있다.

ABSTRACT

As industrial information disclosure by insider's rate is around 80%, most of relevant researches explain briefly its causes are discontent of salary or human resources system. This paper scrapes texts on Jobplanet, an anonymous company review website and analyzes discontent keyword by 7 related area and their contexts to find out more details on brief causes referred above.

After drawing LGG (Local Grammar Graph) by each areas with related dictionary list, this paper shows an example of concordance as a proof and several ways for human resources leakage prevention. Finally, text analysis results are compared with previous researches based on survey with limited questions and answers. This study is meaningful to expand the scope of employee discontent analysis with company review text and provide more specific, granular and honest discontent vocabularies.

키워드 : 텍스트 분석, 기업 리뷰, 텍스트마이닝, 직원 불만, 부분문법그래프

Keywords : Text Analysis, Company Review, Text Mining, Employee Discontent, LGG

Received 26 February 2019, Revised 28 February 2019, Accepted 27 March 2019

*Corresponding Author Yongsuk Park (E-mail:yongspark@sjcu.ac.kr, Tel:+82-2-2204-8062)

Professor, Graduate School of Information Security, Sejong Cyber University, Seoul, 05000 Korea

본 논문은 해당논문 저자 백혜연의 석사학위 논문을 기반으로 하였음.

Open Access <http://doi.org/10.6109/jkiice.2019.23.4.357>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

기업 정보 유출, 산업기밀 유출등과 같은 산업스파이 범죄는 해를 거듭할수록 증가하는 추세이다. 국가정보원 산업기밀보호센터의 자료에 따르면 전현직 직원에 의한 내부유출이 80%를 차지하며 협력업체 직원의 숫자까지 더하면 기술유출 주체의 90%가 관련 업무를 하는 직원이라고 한다. 특히 산업보안에 의한 의식이 대기업보다 낮은 중소기업에서의 기술유출이 지속적으로 증가하고 있다.[1] 기술유출의 원인이 금전유혹, 개인영리, 인사 및 처우 불만으로만 조사되었기에[1] 더 구체적인 불만 내용을 확인하는 것이 본 연구의 첫 번째 목적이다. 해당 불만을 해소하거나 최소화시킴으로써 직원이 기업에 대해 가지는 불만으로 인한 정보 유출 횟수도 줄어들 가능성이 있다. 그리고 기업에 대한 불만이 반드시 정보 유출(정보 절도)로 이어지지는 않더라도 기업 불만을 해소함으로써 퇴사나 이직으로 인한 인력 유출을 방지하는 효과를 얻을 수 있다.

본 연구에서는 익명기업리뷰를 분석하여 전현직 직원들이 기업에 대해 가지는 불만을 구체적으로 분석하여 기존 연구에 나타난 이직 및 정보유출 원인과 비교한다. 일반적으로 텍스트 분석 방법과 관련한 기존 연구들은 문서정보추출이나 오피니언 자동 분류를 위한 연구에 부분문법그래프, LGG(Local Grammar Graph) [2], [3]를 사용하는 것이 한 가지 방법이다. 영화평 문서의 오피니언 분류 연구[2]에서는 제한된 언어적 정보를 바탕으로 대다수의 오피니언 마이닝 연구들이 주로 통계에 의해 수행되고 있던 상황에서 어휘, 통사, 의미적 언어 기술을 통해 더욱 향상된 성과를 보였다. 웹 문서 정보추출과 자연어처리 관련 연구[3]에서는 웹 문서 정보추출을 위해 웹 텍스트를 검토하여 자연어 문장들을 추출하였다. 신문사의 주식의 등락과 관련한 뉴스 기사를 기반으로 LGG를 구축하여 주식정보 등락과 관련한 영역의 의미지식 데이터베이스를 자동으로 구축하였다.

본 연구에서는 기업리뷰 텍스트 분석 결과를 바탕으로 기업 불만에 관한 사전 목록 및 LGG를 구축하여 구체적인 불만사항들을 파악하였다. 뿐만 아니라 콘코던스(용례 색인, Concordance)로 LGG의 기능 및 실제 결과를 확인하고 이를 바탕으로 기업에 대한 불만해소를 통한 개선방안을 제시하려 한다. 이는 주로 설문이나 문헌 조사를 통해 진행되었던 기존 연구와 다른 점이다.

또한 세부 의미 분야별 의미사전을 구축하고 동사, 형용사 기본사전을 통해 활용형을 포함함으로써 다양한 어휘 변화 가능성을 고려하였다. 추후 의미사전 목록을 추가하여 더욱 다양한 불만 표현을 더 효과적으로 분석할 수 있을 것이다. 특히 기존 텍스트마이닝 연구들[4], [5],[6]에서 명사 위주의 사전을 제시했다면 본 연구에서는 의미사전 목록에 명사뿐만 아니라 서술어를 함께 제시하여 더욱 확실한 불만 내용 파악이 가능하게 되었다.

산업정보유출 방지와 관련된 기존의 연구에는 인력 관리를 통한 방지 방안으로는 정보유출 행위자와 행위의 원인에 대한 연구[7],[8]와 정보유출을 직원절도행위 중 정보절도로 정의한 연구[9]가 있다. 정보절도를 정보유출행위로 정의하고 그 원인을 개인적, 조직적, 환경적 요인으로 나누어 분석하였다[9]. 특히 소속 기업불만과 관련하여 인원보안의 중요성에 대한 연구[10]가 있었다. 그리고 설문을 통한 직원들의 퇴사 원인 연구[10], [11]들이 있었다. 퇴사한 직원들의 이직원인 연구[10]에서는 임금이나 직무 혹은 승진, 감독자 및 동료에 대한 만족도가 낮거나 불만이 발생할 경우 이직이 발생할 수 있다고 하였다. 신입사원 대상의 설문을 바탕으로 한 조기퇴사 원인 연구[11]에서는 근로여건 (근로시간, 보수, 승진, 장애전망, 근로환경 등) 불만족이 빠른 퇴사의 첫 번째 원인이었다.

앞의 연구들을 통해 산업정보유출의 원인과 이직 및 퇴사의 원인이 유사하다는 것을 확인하였다. 다만 설문을 통해 이루어진 연구이기에 설문 항목이나 답변에 나온 항목들만이 연구 결과로 수집되는 한계가 있다. 본 연구에서는 이 연구들을 바탕으로 전체적인 이직 및 퇴사 원인 유형을 구분하고 텍스트 분석을 통해 얻은 구체적인 다양한 결과를 기존 연구들과 비교하여 산업인력유출 원인을 도출하고 유출 방지 방안을 모색한다.

II. 연구방법 - 기업 내부자 불만 분석

본 연구에서는 우선 익명 기업리뷰 텍스트를 분석하고 이 결과를 바탕으로 기업불만에 관한 의미 사전을 구성한다. 다음으로 LGG를 통해 구체적인 불만사항들의 패턴을 구축한다. 마지막으로 콘코던스를 통해 LGG의 기능을 검증하고 실제 결과를 바탕으로 기업에 대한 불

만내용을 확인한다. 이를 통해 기존 설문을 통해 진행되었던 연구보다 더욱 구체적인 결과를 얻을 수 있다. 총 7개의 세부 의미 분야별 의미사전을 구축하고 동사, 형용사의 활용형을 포함함으로써 다양한 어휘 변화 가능성을 고려한다.

2.1. 익명 기업 평가 웹사이트에 언급된 전현직 직원의 기업 리뷰 텍스트 수집

특정 기업에 대한 리뷰가 아닌 가능한 불특정 다수의 기업에 대한 직원들의 생각을 파악하기 위한 목적으로 텍스트를 수집하였다. 익명 기업 리뷰 웹사이트인 잡플래닛(www.jobplanet.co.kr)에 2018년 7월 21일 기준 총 19189개 기업의 기업리뷰가 등록되어 있었고 1개 기업당 1건씩의 기업리뷰를 수집하였다. 특히 실제 근무 경험이 확실하게 드러나기에 일반 상품 후기나 댓글, SNS 텍스트보다 내용에 지나치게 감정적이거나 불확실한 질문, 허구적인 내용이 적다고 판단하여 잡플래닛의 기업 리뷰 텍스트를 분석 대상으로 선정하였다. 잡플래닛의 기업 리뷰는 전체 리뷰, 장점, 단점으로 이루어져 있는데 이 중 전체 리뷰(따옴표 내부) 부분만을 크롤링하여 텍스트 분석에 사용하였다. 장점 및 단점 하위 텍스트의 경우 이미 긍정적, 부정적 특성이 드러나기에 가능한 포괄적인 내용을 포함한 전체리뷰 부분을 선택하였다. 수집한 부분은 [그림1]에서 빨간색으로 표시된 부분이다.

파이썬 3.7.0 버전에서 BeautifulSoup 패키지를 사용 [12],[13]하여 약 2만 건의 기업 후기 (1기업 당 후기 1건, 총 19189건) 텍스트를 수집하였다.

★★★★★ 5.0

IT/웹/통신 | 4기업리뷰

"직원 모두가 대우받으며 즐겁게 일할 수 있는 곳입니다."

장점

수평적인 사무실 분위기, 안정적인 연봉, 복리후생이 잘되어있음, 업무시

단점

연봉은 충분하지만, 업무량이 의외로 많았다는 점밖에 없음.

Fig. 1 Crawling part of Jobplanet review (captured)

2.2. 유니텍스 프로그램을 통한 텍스트 분석

유니텍스(Unitex)는 사전과 문법 기반의 다국어 코퍼

스 분석 프로그램으로써 프랑스 파리에 있는 마르라발레 (Paris-Est Marne-la-Vallée) 대학교의 LIGM 연구소에서 개발되었고 오픈소스(<http://unitexgramlab.org/>)로 공개되어 있으며 다양한 플랫폼에서 사용이 가능하다 [14]. 약 22개의 언어를 지원하며 사용하려는 언어별 사전을 적용하여 코퍼스를 분석할 수 있다. 코퍼스를 구축하는 데에는 파이썬(Python)을 사용하였으나 파이썬용 한국어 정보처리 패키지 KoNLPy (Korean Natural Language Processing in Python)를 사용하지 않은 이유는 해당 패키지에 포함된 한국어 사전의 품질 때문이다. 복합 명사가 과도하게 분리되어 [15] 원래의 의미를 찾기 힘들어지거나 복합 동사나 서술어의 원형이 잘못 추출될 경우 전체적인 빈도 계산이나 의미 분야별 어휘 추출 작업에 오히려 혼란이 생길 수도 있다고 생각하여 코퍼스에 사용된 텍스트를 그대로 관찰해보기로 했다. 그리고 자유롭게 작성되는 온라인 후기의 특성상 띄어쓰기가 무시되고 새로운 유형의 복합어들이 많이 생성되므로 그런 특성을 가진 어휘들 자체를 확인할 필요가 있었다. 한국어사전이 적용되지 않은 기본 상태의 유니텍스 프로그램은 띄어쓰기가 되지 않은 한국어 단어를 제대로 구분하지 못하였는데 오히려 유니텍스를 통하여 띄어쓰기 기준으로만 분리된 모든 단어들을 확인할 수 있었다. 단어들을 의미 분야별로 분류하고 유니텍스 그래프 모듈을 통하여 LGG[16]를 구축하였다. 전체적으로 유사한 의미를 가지는 키워드들을 추출하여, 의미, 행위 분야별 키워드들을 분류하고 불만 관련 문장 패턴을 구축하는 연구에 중점을 두었다.

2.3. 키워드 유형 분류 및 의미 사전 구축

기업리뷰에 등장한 키워드들을 구분하여 의미 분야, 주제, 행위별 유니텍스용 의미 사전을 구축하였다. 복합어와 복합어절의 경우 텍스트 본문에 나타난 대로 띄어쓰기가 무시된 내용은 무시된 채로 기재하였다. 개별 어휘들의 빈도보다 다양한 키워드 유형(type)을 추출하는 작업에 중점을 두었다. 텍스트 마이닝을 이용한 청소년 문제 토픽 모델링 연구[4]에서는 토픽 선정을 위해 뉴스 기사 텍스트를 분석했는데 한글 보통명사들만을 추출하여 관련 서술어에 의해 전체 문장이나 문단의 의미가 변경되는 가능성을 고려하지 못하였다. 빅데이터 기반 정보 추천 시스템 연구[5]에서는 의료관련 텍스트 데이터 분석을 통해 질환의료키워드사전을 구축하였는데

질환명과 함께 사용되는 서술어 사전도 포함된다면 더욱 정확한 질환 관련 정보 검색과 추천이 가능할 것이다. 또한 영문 의미 기업리뷰 사이트 텍스트를 분석하여 직원 만족도와 기업 성과의 연관성을 분석한 연구[6]에서도 명사로만 이루어진 의미 분야별 키워드를 제시하였다. 본 연구에서는 의미사전 목록에 명사와 서술어를 포함하여 더욱 정확한 내용 파악이 가능한 언어 자원을 구축하려 한다.

III. 익명 기업 리뷰 텍스트 분석 결과

3.1. 전현직 직원의 퇴사(인력유출) 원인

텍스트 분석 결과에서 각종 기호 및 ‘하다’ 등의 기본 서술어 등을 제외한 유의미한 키워드들을 살펴보았다. 500건 이상의 고빈도어들은 사람(직원) 3804건 <19.4%>, 없다(부재하다, 전무하다, 부족하다, 제로) 3393건 <17.3%>, 분위기(문화) 2635건 <13.4%>, 발전(성장, 미래, 전망) 1924건 <9.8%>, 연봉(월급) 1382건 <7.1%>, 배우다(배울 점) 1281건 <6.5%>, 추천(비추천 포함) 1157건 <5.9%>, 복지(복리) 1100건 <5.6%>, 체계(시스템) 1083건 <5.5%>, 야근(야간) 655건 <3.3%>, 환경(공간) 647건 <3.3%>, 퇴사(이직, 사직) 535건 <2.7%>의 순서로 나타났다.

‘추천’의 경우 ‘추천하지 않는다’는 의미로 많이 사용되었고 “야근이나 박봉도 견딜 수 있는 분에게는 (이 회사 임사를) 추천합니다.”와 같이 반어법으로 사용된 경우도 다수였다. 그러므로 ‘추천하다’라는 키워드 자체가 많다는 것만으로 추천받는 기업이 다수라고 할 수 없었다.

우선 상위 고빈도어 중에서 유일한 서술어인 ‘부재하다/없다’의 대상들을 세부 의미 분야별로 구분하여 표1에 기재하였다. 각각의 어휘들이 부재한다고 생각해보면 앞으로 살펴볼 7개의 의미 분야별 불만 내용과 큰 차이가 없다는 것을 확인할 수 있다. 잡플래닛에 등록된 익명기업리뷰가 보통의 상품 및 서비스 후기들과 다른 점은 싫다, 나쁘다, 불편/불쾌하다 등의 불만을 나타내는 직접적인 부정어휘 없이 ‘부재’한다는 사실 전달만으로 부정적 정보를 나타내는 텍스트들이 다수 존재한다는 것이다. 표1을 보면 체계, 인력, 복지, 보상, 휴가가 없는 것뿐만 아니라 ‘배려와 인정’이 없고 ‘전문성과 기

술력’이 없으며 ‘배울 점’과 미래가 없음으로써 기업에 대한 불만이 발생하게 된다는 것을 알 수 있다. 각기 다른 기업에 대해 서로 다른 전현직 직원들이 ‘부재한다’는 사실만으로 유사한 내용의 부정적 정보를 표현하고 있다.

Table. 1 Examples of Negative Vocabularies by Absence

Area	Vocabulary
System	system, process, culture of organization
Human	leader, human resource, person in long service
Environment	resting room, window, safety equipment
Rest	work life balance, holiday, vacance, weekend
Benefit	welfare, benefit, education, supporting
Salary	pay, salary, bonus, allowance, reward
Mood	future, recognition, care, regard, consideration
Work	professionalism, task, mission, competitiveness

다음은 상위 (500건 이상) 고빈도어 순서로 유사의미를 가지거나 동일한 분야에서 사용될 수 있는 어휘들을 정리한 것이다. 인물, 분위기(문화), 연봉, 복지, 야간근무(야근), 물리적 근무환경(공간), 인사(퇴사, 이직, 사직) 총 7개 분야의 유형으로 구분하였다. 고빈도어 중 ‘발전’과 ‘배우다’의 경우 이미 언급한 ‘성장, 미래, 전망’ 및 ‘배울 점’ 외의 세부 키워드 유형은 찾을 수가 없어서 따로 소제목으로 분류하지 않았다. 각 유형의 개별 어휘들은 ‘존재’하거나 언급됨으로써 부정적인 사실을 나타내는 유형과 부정 서술어나 ‘부재’함으로써 부정을 나타내는 유형으로 구분되었다.

3.1.1. 인물 관련 - 사람(직원) 19.4%

경영진 (중립-CEO 간부 경영진 고용주 고용주 과장 오너 팀장, 부정-독재자 윗놈 윗대가리 약덕대표), 그 외 인물 (중립-사원 인력 직원 사람 가족, 부정- 끈대 꼴통 낙하산 똥차 빌런 양아치) 등으로 구분하였다.

기업 및 경영진의 태도와 관련하여 ‘취급’, ‘대접’, ‘대우’에 해당하는 키워드들(소모품 <101건> 부속품 가족)을 사전 목록으로 구축하였다. 긍정 어휘(예우 배려 상호존중 상호존칭 존대 존칭)의 경우 ‘없다’와 호응하여 ‘(직원에 대한) 배려가 없다’는 식의 불만으로 연결되었다. ‘차별, 불평등, 무시(하다)’와 같이 그 자체로 불만 극성을 가지는 어휘들도 다수였다. 정규직(비정규직 포

함)은 168건으로 다소 낮았는데 정규직과 비정규직의 구분이 없는 중소기업에 대한 리뷰 개수가 높았기 때문에 나타난 결과로 보인다.

3.1.2. 분위기, 문화 관련 - 분위기(문화) 13.4%

전반적인 분위기와 관련하여 그 자체로 불만인 내용(가부장적 강압적 강제적 경직적 고압적 편파적 권위주의 보수적 봉건적 비체계적 전근대적 군대식 남성중심 주먹구구식)과 중립 및 긍정적(애착 인간적 능력인정 활기 생기 성취감 연대감 존중) 내용들이 확인되었다. ‘동료애’와 같은 어휘는 ‘있다/느끼다/크다’ 등과 함께 사용되어 긍정을 나타내기도 하지만 ‘없다/부재하다’ 등과 함께 사용되면 불만을 나타내므로 긍정적 어휘는 서술어와 함께 실제 내용을 파악해야 했다.

3.1.3. 임금 관련 - 연봉(월급) 7.1%

임금 관련 다양한 표현(급여 연봉 월급 기본급 보수 임금)이 나타났고 기타급여(보너스 수당 경비 복지비) 관련 표현들이 확인되었다. 급여 및 연봉이 적다(낮다 작다 짜다 박하다)는 언급과 임금이 밀린다, 혹은 체불된다(안 나오다 떼이다 못 받다)는 내용도 다수 있었다.

3.1.4. 복지 및 휴가 관련 - 복지(복리) 5.6%

복지제도(복지 건강검진 교육 리프레쉬 복리 복리후생 혜택) 및 휴가(휴무 월차 연차 반차 휴식)관련 어휘들이 추출되었다. 어휘들 자체에는 부정 극성이 없지만 ‘없다’, ‘사용하지 못하다’ 등의 서술어와 호응하여 부정적 사실을 나타내는 리뷰들을 다수 확인하였다.

3.1.5. 근무시간 관련 - 야근(야간 근무) 3.3%

근무시간과 관련한 표현(워라밸 근무시간 업무시간 여가생활 자기생활 개인시간). 그 자체로는 중립이지만 ‘워라밸이 지켜지지 않는다’, ‘근무시간이 길다’, ‘업무량이 많다’와 같이 부정적 언급의 대상이 되었다. 그리고 ‘강제 야근’과 같이 강제적으로 이루어지는 추가 근무에 대한 불만도 다수였다.

3.1.6. 근무환경 관련 - 환경(공간) 3.3%

물리적인 근무환경 중에서 공간(근무시설 구내식당 휴게시설 휴식공간 기숙사 사무실), 환경(업무환경 간접흡연 한여름 공기 추위 햇볕), 위치(교통편 근무위치

시골 외지), 불편함(덥다 춥다 더럽다 허름하다 춥다)등으로 구분하였다. ‘휴식공간이 없다’와 같이 부재하여 불만이 되는 항목도 있었다. 기존 설문을 통한 연구에서는 자세하게 확인할 수 없었던 내용이다.

3.1.7. 채용, 인사 관련 - 퇴사(이직, 사직) 2.7%

채용 및 인사와 관련하여 기업의 부정적 행위(해고하다 퇴사시키다 자르다 내보내다 갈아치우다 노동착취), 직원의 부정적 행위(퇴사 <295건> 이직 <219건> 사직 <21건> 나가다 그만두다 떠나다 관두다), 직원의 긍정적 행위(오래근무 근속하다 진급하다), 고용관련 부정(인력난 인사적체 고용불안정 좌천 자리싸움)등으로 분류되었다. 쉽게 해고하고 해고당하는(권고사직당하다 나가야하다 아웃당하다), 쉽게 이직하거나 인력이 부족한 상황임에도 신규 채용을 하지 않는 분위기에 대한 언급이 많았다.

3.1.8. 소론

전현직 직원들이 기업에 대해 가지는 불만과 관련하여 7가지 의미 분야별 사전 목록 일부를 제시하였다. 부정적 어휘가 ‘없다(안하다)’와 결합하여 ‘야근 안하다, 차별이 없다’와 같은 긍정적 정보를 나타내기도 하였지만 긍정적 어휘가 ‘없다(부재하다)’와 함께 사용되어 부정적 정보를 나타내는 경우가 다수였다. 결국 기존 연구 [6]에서 설문을 통해 첫 번째 퇴사 사유로 선정된 “근로여건(근로시간, 보수, 승진, 장래전망, 근로환경 등) 불만족” 및 산업정보유출의 원인 연구[2],[4]에서 언급한 “처우 불만”, “인사 불만”과 같은 내용인 것이다. 텍스트 분석을 통해 기존 연구에 나타난 직원 불만의 세부 내용을 더욱 구체적으로 확인할 수 있었다.

또한 소수이지만 범죄와 관련된 어휘들(비리, 성범죄, 횡령, 탈세 등)도 나타났다. 빈도가 10 이하로 낮아서 따로 유형을 분류하지는 않았으나 범죄 관련 어휘들의 경우 언급 자체만으로 부정의 정도가 상당히 높아질 수 있다. 기존 설문을 통한 연구에서는 확인할 수 없었던 의미 분야이다.

‘이러저러한 이유로 퇴사하였다, 퇴사할 예정이다’라고 분명하게 언급한 리뷰는 거의 전무하였다. 다만 2개 분야의 불만이 중첩되거나 한 가지 분야의 불만 정도가 클 경우 퇴사나 이직을 결심하는 동기가 될 수 있을 것이다.

3.2. 기업 불만 LGG (Local Grammar Graph) 구축

LGG 방법론은 유한문법(Finite Grammar)모델로 정규문법과 문맥자유문법 사이의 표현력을 가지는 모델로써 프랑스의 전산언어학자인 모리스 그로스(Maurice Gross)가 제안[2]하였다. 왼쪽에서 오른쪽으로 진행되는 방향성 그래프(Directed Acyclic Graph) 형태로 구현되며 LGG 문법을 통해 제한된 어휘적 현상이나 일반 통사규칙으로 설명하기 어려운 관용적 표현들을 효율적으로 기술할 수 있다. 유니텍스 프로그램의 그래프 모듈에서 생성 후 유한상태 오토마타 (Finite-State Automata)로 자동으로 컴파일되어[2] 실제 텍스트 분석 작업에 적용이 가능하다.

앞에서 정리한 의미 분야별 사전목록을 바탕으로 LGG를 구축하였다. 지면의 한계로 ‘대표 및 경영진 관련 불만LGG’ 1개만 제시한다. 전체 LGG는 직원 불만과 관련한 텍스트 마이닝 연구에서 확인할 수 있다.[12]

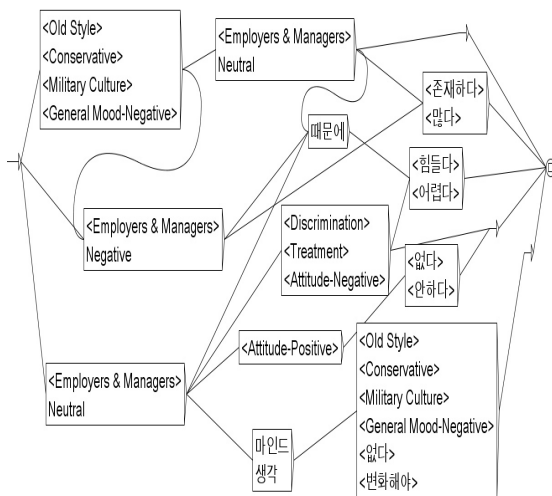


Fig. 2 LGG for CEO & Managers Discontent

대표 및 경영진 관련 불만 LGG(그림 2)에서는 중립적인 키워드들은 부정적인 서술어와 호응하여 부정적인 내용으로, 부정적인 키워드들은 언급된 자체만으로 부정적인 서술어 되는 것을 확인할 수 있었다. 그리고 서술어뿐만 아니라 ‘구시대적’, ‘군대식’, ‘강압적’, ‘수직적’과 같이 전반적인 분위기 서술(<General Mood-Negative>)에 나타난 사전 목록들과 호응하여 부정적인 정보를 나타내기도 하였다.

3.3. 기업 불만 콘코던스 확인

다음은 대표 및 경영진 관련 불만 LGG에 대한 콘코던스 예시이다. 결과 일부를 캡처한 그림 3의 파란 색 밑줄 앞뒤 내용을 살펴보면 LGG를 통한 의미 분야별 불만 내용의 결과를 확인할 수 있다.

그때 그때만 넘기는 성향 회사 대표이사 마음대로 하는 회사, 어디나 같을 수 있지만 유지보수를 배울 수 있음 대표이사 중심으로 운영되고있는 회사이며, 기본적으로 디해진 사규 없이 그때 그때 사장 기본에 따라 모든것이 결정되는 곳 신입 뽑으면 한 두 회사 생활을 할 수 있다.(S) 사장 눈치 보는 회사임, 사장 맘에 들면 만사가 오케이이고 클레베이터도 없음.(S) 사장 눈치 보느라 알아서 벌벌.(S) 사장은 뻑하면 내가 사장해 3등할 군대식 문화.(S) 사장 눈치 보고, 누가 뭐라고 안해도 알아서 눈치 보는 문화.(S) 거 안 나옴.(S) 모든 게 사장 마음대로.(S) 유통계 디자인을 경험하기 좋음 배우는게 ! 여행사라서 그런지 모든게 사장 마음대로 움직이는 시스템.(S) 함 도전하고픈 회사. 2번째 하는 회사.(S) 워든 사장 마음대로.(S) 8시반 출근인데 사장이 보는 자료를 8시 친절해야함 체계없음.(S) 워든지 사장 마음대로.(S) 연차없음 다방면의 사업확장으 3형 소기업 요즘 젊은이들이 사장 비위 맞춰주며 다니기엔 오래 다닐 수 없을듯 가족 는 세계최초의 연구소임.(S) 사장 중심으로 돌아가는 회사모든 분야에 사장이 관여 6 곳.(S) 회사가 너무 사장 중심으로 돌아간다.(S) 연봉이 빨리 오른다.(S) 다양한 사업 1 직원들은 편하지만 몇명의 상급자 때문에 분위기가 별로임 관계속 깊은 신뢰의 워드

Fig. 3 Concordance Example for CEO Discontent LGG

기존 연구들에서는 퇴사나 이직 및 산업정보 유출의 원인이 첫째로 금전적인 문제(낮은 보수 및 적은 금전적 보상, 처우불만)에 있다고 하였는데[1],[10] 그뿐만 아니라 임직원들의 직원들을 대하는 태도, 기업의 분위기도 직원 불만의 상당 부분을 차지하며 퇴사나 이직의 원인이 될 수 있다는 것을 텍스트 분석을 통해 알 수 있었다.

IV. 전현직 직원의 불만 해소 및 퇴사(인력 유출) 방지 방안

직접적으로 산업정보유출사고가 발생하지 않더라도 인력이 유출됨으로써 장기간에 걸쳐 정보유출과 같은 결과가 나타날 수 있다. 결국 인력이 유출되지 않도록 노력하는 것이 기업의 산업정보보호는 물론이고 장기적으로는 채용비용 감소 등 기업에 이익이 될 수 있다.

전현직 직원들은 기업의 분위기에 대해 언급을 많이 하였다. 부정적인 분위기(군대문화, 끈대문화, 강압적, 강제적, 상하수직 관계)를 개선하여 대표를 비롯한 경영진들이 일반직원들을 대하는 자세를 바꾸고 일반직원들과 의사소통을 제대로 할 수 있다면 직원들의 퇴사 의지가 감소할 수 있다. 그리고 체계가 없고 전망(미래, 발전)이 없다는 불만과 관련하여 전사적으로 업무 체계를

갖추고 더욱 성장할 수 있다는 미래비전을 직원들에게 제시해야 할 것이다. 불필요한 야간 근무나 휴일 근무를 없애고 추가 근무를 하게 될 경우 그에 대한 수당 지급도 당연하다. 복지 체계가 없거나 사용이 힘든 기업은 직원을 위한 실질적인 복지제도를 만들 필요가 있다. ‘배울 점이 없다’는 불만과 관련하여서는 직무 교육이나 기타 교육 프로그램 운영뿐만 아니라 사람을 통해 배울 수 있는 시스템과 문화를 만들 필요가 있다. 결국 CEO를 비롯한 경영진들이 직원들을 함부로 대하는 태도를 바꾸고 인력을 존중해주는 분위기를 만들어 나간다면 일반 직원들도 그러한 분위기와 문화를 체득하게 될 것이다. 여러 방면으로 직원들이 자발적인 ‘애사심’을 갖도록 하는 것이 산업 인력 및 산업 정보 유출 방지 결과로 연결될 수 있다.

V. 결 론

본 연구에서는 전현직 직원들이 직접 작성한 익명 기업리뷰 텍스트를 분석하여 기존의 설문을 통한 연구에서 ‘보수 불만, 처우 불만, 승진 누락’ 등으로만 분석된 인력 및 산업 정보 유출의 원인을 더욱 구체적으로 파악하였다.

예를 들어, ‘사람을 중요시하는’, ‘직원을 쉽게 해고하지 않는’, ‘다양한 교육과 복지가 이루어지는’, ‘적절한 기본급 및 수당, 복지비 지급이 이루어지는’, ‘미래 비전이 있는’, ‘경영진들이 제대로 업무를 하는’ 분위기의 기업으로 바뀌고 ‘정보보호 관련 체계가 있는’ 기업이 된다면 퇴사율도 감소하고 더불어 내부인력의 자발적인 이직이나 스카우트에 의한 산업 정보유출의 위험도 낮아질 것이다.

이는 기존 설문조사를 기반으로 한 연구 대비 웹사이트 등에 공개된 온라인 텍스트 분석을 통해 좀 더 다양하고 자세한 불만 내용과 퇴사 원인을 분석할 수 있다는 가능성을 보여주었다. 구체적으로는 의미사전과 LGG, 콘코던스를 모두 사용함으로써 텍스트마이닝 방법의 완성도를 높였다고 할 수 있다.

이번 분석에 사용한 익명 기업리뷰 텍스트들이 100% 사실이 아닐 수도 있다는 한계점은 존재한다. 예를 들어 일부 비자발적으로 이직한 직원의 경우 직원 자체의 태도나 업무능력에 문제가 있는데도 기업 탓만 하는 식의

리뷰가 작성되었을 가능성도 있는 것이다. 또한 회사와 경영진들의 입장에서는 전현직 직원들에 대해 또 다른 차원의 불만을 가지고 있을 수 있다. 추후 CEO 및 경영진들의 불만사항에 대해 분석을 하거나 직원들이 기업에 대해 만족하는 내용들에 대한 텍스트 분석 작업과 같은 추가적인 연구가 요구된다. 이를 통해 결과 비교에 의한 상호 간의 불만 내용을 더욱 구체적으로 파악하여 개선 방안을 마련할 수 있을 것이다.

또한 교육 프로그램 운영과 관련하여 직원뿐만 아니라 임원진들을 대상으로 하는 교육도 필요하며 해당 교육의 내용과 관련한 추가적인 연구가 이루어진다면 더욱 효과적인 산업인력 유출 방지 방안이 될 것이다.

마지막으로 ‘군대문화, 군대분위기, 가족기업, 가족회사’와 같이 어휘 그 자체로는 극성(긍정이나 부정)을 가지지 않는 표현이 특정 의미 분야(기업 리뷰 텍스트)에 사용됨으로써 부정 극성을 나타내거나 반어법을 사용하여 부정적 의견을 나타내는 경우도 발견되었다. 그 자체로 확실한 극성을 나타내는 어휘뿐만 아니라 의미 분야에 따라 극성이 달라지거나 존재하지 않았던 극성이 생겨나는 표현들에 대한 추가 연구가 이루어진다면 더욱 정교한 텍스트 분석이 가능해질 것이다.

REFERENCES

- [1] NIS Industrial Security Protect Center. Title. Industrial Espionage Disclosure Statistics Information graphic, 2014 [Internet] Available: <http://www.etnews.com/20141123000016>.
- [2] J. S. Nam. “Study on Linguistic Patterns of Online Reviews on Movie for the Automatic Classification of Human Opinion,” *The Linguistic Society of Korea*, vol. 58, pp. 75-103, 2010.
- [3] M. K. Kim, and Y.W. Lee. “A Study of Automatic Ontology Building by Web Information Extraction and Natural Language Processing,” *The Journal of the Institute of Internet, Broadcasting and Communication*, vol. 9, no. 3, pp. 61-67, 2009.
- [4] J. Y. Cho, and K. W. Cho. “Topic Modeling on the Adolescent Problem Using Text Mining,” *Journal of Korea Institute of Information and Communication Engineering*, vol. 22, no. 12, pp. 1589-1595, 2018.
- [5] J. C. Lee, and M. H. Lee. “Big data-based information recommendation system,” *Journal of Korea Institute of*

- Information and Communication Engineering*, vol. 22, no. 3, pp. 443-450, 2018.
- [6] N. Luo, Y. Zhou, and John J. Shon, "Employee Satisfaction and Corporate Performance: Mining Employee Reviews on Glassdoor.com," in *37th International Conference on Information Systems*, Dublin, pp. 387-402, 2016.
- [7] H. M. Yeon, "Strategy for prevention of Trade Secret Leaks by Insiders," Thesis of Graduate School of Strategic Studies, SungKyunKwan University, 2013.
- [8] T. G. Lee, "Prevention of Industrial Information Leakage & Methods for Managing Personnel Security," Thesis of Graduate School of Strategic Studies, SungKyunKwan University, 2011.
- [9] P. S. Kim. "A Study on Influential Factors for Employee Theft," Doctoral dissertation of Graduate School of KwangWoon University, 2015.
- [10] T. H. Kim, "The analysis of Employee's turnover factors through cause analysis of Behavior Engineering Models in HPT," Thesis of Graduate School of HRD, Chung-Ang University, 2008.
- [11] D. W. Eom, "The Status and Causes of Early Separation of College Graduate Newcomers: Focusing on the HRM Perspectives," *Journal of Vocational Education & Training*, vol. 11, no. 2, pp. 237-260, 2008.
- [12] H. Y. Baek, "Prevention of human resources leakage and industrial information disclosure by text mining on staffs discontent," Thesis of Graduate school of Information Security, Sejong Cyber University. (Under preparation, expected August 2019.)
- [13] J. T. Park. *Web Crawling learning by Python*, Seoul: Information Publishing Group, 2018.
- [14] UNITEX [Internet] Available: <https://unitexgramlab.org/>
- [15] KoNLPy Morphme analysis and part of speech tagging [Internet] Available: <http://konlpy.org/ko/latest/morph/>
- [16] UNITEX User Manual [Internet] Available: <https://unitexgramlab.org/releases/3.1/man/Unitex-GramLab-3.1-usermanual-en.pdf>



백혜연(HyeYeon Baek)

한국외국어대학교 프랑스어과(학사)
한국외국어대학교 언어인지과학과 (석사 수료)
한국방송통신대학교 컴퓨터과학과(학사)
세종사이버대학교 정보보호대학원 석사과정
※ 관심분야 : 텍스트 분석, 한국어전자사전, 산업보안, 정보추출 등



박용석(Yongsuk Park)

서강대학교 컴퓨터학 (학사)
뉴욕(POLY) 대 (석사, 박사)
AT&T (Bell) Labs, 삼성전자
현재 세종사이버대학교 정보보호대학원 주임교수
현재 세종사이버대학교 IT학부 교수
※ 관심분야 : IT서비스 및 보안, 산업보안, 4차산업혁명, 클라우드, IoT 등