

채용정보 분석을 통한 비즈니스 직무 스펙 연구*

이중화** · 이현규***

〈목 차〉

- | | |
|---------------|------------------|
| I. 서론 | III. 연구방법과 프레임워크 |
| II. 이론적 배경 | IV. 실험과 연구결과 |
| 2.1 채용 트렌드 분석 | V. 결론 및 향후 연구과제 |
| 2.2 TF-IDF | 참고문헌 |
| 2.3 NMF | <Abstract> |

I. 서론

2020년은 COVID-19 대유행으로 전 세계인의 삶을 크게 혼란에 빠뜨렸던 해이다. 서비스 산업의 일자리는 큰 충격을 받았으며 디지털 플랫폼의 의존도를 높이고 있다. 1990년대 초의 정보통신 기술(ICT) 혁명은 인터넷의 급속한 확산으로 여러 경제 부문을 변화시키고 지역, 국가 및 국제 시장을 재편시키고 있다. 기업은 상대적으로 저렴한 비용의 글로벌 공급망을 통해 하도급, 아웃소싱 등의 경영전략을 통하여 지리적인 문제를 해결하였다. 광대역 연결의 확장고속 인터넷의 가용성은 2000년대 초부터 디지털 인프라의 급속한 발전을 가능하게 했다. 기업과 개인 모두가 인터넷과 ICT 장치를 널리

사용함으로써 웹 기반 플랫폼(Amazon, eBay et al.)의 길을 열었고 디지털 경제의 토대가 마련되었다(Castells, 2010).

지난 10년 동안 클라우드 인프라와 컴퓨팅 서비스의 유용성은 경제의 모든 부문에 점진적으로 침투한 디지털 플랫폼의 성장을 촉진했다. 일반적으로 플랫폼은 크게 ‘제품 플랫폼’, ‘SW 플랫폼’, ‘서비스 플랫폼’으로 구분할 수 있다. Surie and Sharma(2019)는 서비스 플랫폼은 세 가지 범주로 구분을 하였다(Surie and Sharma, 2019). 소셜 미디어와 같은 개별 사용자에게 디지털 서비스 및 제품을 제공하는 플랫폼, 전자상거래 또는 B2B (Business-to-Business) 플랫폼과 같은 상품 및 서비스의 교환을 중재하는 플랫폼, Upwork 또는 Uber와 같은 디지털 노동 플랫폼을 포함하여 기업, 근로자 및 소비자

* 이 논문은 부경대학교 자율창의학술연구비(2021년)에 의하여 연구되었음.

** 동의대학교 e비즈니스학과, jhlee6050@deu.ac.kr(주저자)

*** 부경대학교 경영학부, hyunqlee@pknu.ac.kr(교신저자)

와 같은 다른 사용자 간의 노동 교환을 중재하고 촉진하는 플랫폼으로 구분된다. 디지털 노동 플랫폼은 근로자들에게 소득 창출 기회를 제공하며 여성, 장애인 및 청년과 같은 특정 근로자에게는 유연한 근로 방식이 더 편리할 수 있다. 또한, 근로자의 저임금 또는 계절적 일자리에서 얻은 수입을 보완할 수 있는 길을 열어 주고 있다(Surie and Sharma, 2019).

디지털 노동 플랫폼이 소득 창출 기회에 중요한 원천으로 부상하고 있기에 개발도상국 정부는 디지털 인프라에 투자하고 있으며 일자리를 찾는 이들에게 디지털 기술을 제공하기 위해 민간 부문에서 개발한 교육 프로그램을 지원하기도 한다(Graham et al., 2017; Heeks, 2017).

ILO(International Labour Organization)에 따르면 플랫폼은 두 가지 유형의 작업자에 의존하고 있다. 플랫폼에 직접 고용된 근로자와 플랫폼에 의해 연결되는 근로자로 구분된다. 2010년 스마트폰을 기반으로 한 미국의 자동차 공유 서비스인 Uber의 직접 고용은 26,900명, 연결되는 근로자 5,000,000명으로 조사되었다. 2007년 프리랜서 근로자를 연결해주는 영국 기업인 PeoplePerHour는 50명의 직접 고용근로자와 2,400,000명의 연결된 근로자가 근무하는 것으로 나타났다. 또한, 플랫폼 수의 기하급수적인 증가로 2010년부터 2020년 사이 플랫폼 기업이 5배 증가하였으며 특히, 2020년에는 전년 대비 USA 29%, India 8%, United Kingdom 5% 성장한 것으로 나타났다(ILO, 2021).

팬데믹(Pandemic) 시대 상황에서 민간 기업의 고용 급변화와 플랫폼 기업의 도약, 인공지능으로 인한 일자리 변화 등 고용시장의 변화

는 불가피한 상태이다. 이러한 환경은 더욱 일자리가 필요한 이들에게 이중고를 주고 있다. 팬데믹으로 인한 플랫폼 노동시장 증가와 4차 산업혁명의 도래로 디지털로의 전환이 가속화되는 사회적 현상을 반영한 취업준비생을 위한 채용정보 분석이 절실한 때이다.

본 연구는 급변하는 산업 구조가 반영된 일자리 채용정보 분석을 통하여 직무별 스펙을 제공하고자 한다. 즉, 기업 비즈니스에 필요한 직무 구분과 그 직무에 맞는 우대조건을 구분하여 인덱스를 발굴하고자 한다. 기업 비즈니스에 필요한 직무 스펙을 발굴하기 위해 1) 직무 목록을 선정하고, 2) 채용 포털 웹페이지를 이용하여 채용 정보 URLs를 수집하고 URLs DB를 구축한다. 크롤링 로봇을 이용하여 수집된 직무별 채용정보의 스펙을 스크래핑하며 3) 각 직무별 우대조건을 수집하여 스펙 사전 구축, 4) 같은 직무 내 스펙들의 우선순위 가중치 추출의 TF-IDF(Term Frequency Inverse Document Frequency), 5) 스펙의 특성치를 이용한 감성 차원 축소를 위한 NMF (Non-negative Matrix Factorization) 알고리즘을 적용하고자 한다. 또한, 벡터간의 거리 측정을 위해 6) 코사인 거리(Cosine distance)를 활용하여 비즈니스 직무별 스펙을 개발하고자 한다.

II. 이론적 배경

2.1 채용 트렌드 분석

코로나19는 현대인들의 모든 활동을 제약하고 있다. 특히, 경제 활동의 지표인 개인의 임

금 하락을 초래했다. ILO에 따르면 여성과 저임금 근로자들이 상대적으로 더 많이 일자리를 잃었으며 큰 타격을 받은 것으로 나타나고 있다. 유럽의 여성 근로자는 임금의 8.1%를 잃었으며 남자 근로자의 임금은 5.4%를 잃은 것으로 나타났다. 최저 임금 50%내의 근로자는 임금 17.3%를 잃었으며 다른 계층에 비해 더 취약한 것으로 나타났다(ILO.ORG, 2020).

글로벌 구인 구직 중개업체인 로버트월터스(RobertWalters)는 2020년 고용시장 동향 분석 보고서를 분야별로 발표하였다(Robertwalters, 2020). 먼저, 재무&회계(Accounting&Finance) 분야의 인기 직무는 재무분석가, 재무회계관리자, 세무전문가로 나타났으며 평균 근속 연수는 3~5년으로 나타났다. 또한, 44%가 12개월 이내 이직을 고려하고 있으며 90%가 적극적인 이직 의사가 없더라도 좋은 기회가 있다면 추천받았다고 했다. 이직을 고려하는 이유는 급여, 경력개발, 일과 삶의 균형(Work and Life Balance) 순으로 나타났다. 인사(Human Resource) 분야는 인사 비즈니스 파트너, 개발자 채용 전문가, 인적자원개발 교육 전문가 등이 인기 직무로 나타났으며 2년 미만의 평균 근속 연수를 보이고 있다. 이직을 고려하는 이유로는 경력개발, 조직문화, 임금 순으로 나타났다. 소매&소매(Consumer&Retail) 분야는 e커머스 관리자, 디지털 마케팅 관리자, 면세 채널 관리자 등이 인기 직무이며, 평균 근속 연수는 3~5년이지만 96%가 좋은 기회가 있다면 이직 회사를 추천 희망한 것으로 나타났다. 제조업(Manufacturing) 분야는 반도체 디자인 엔지니어, 자율 주행 시스템 엔지니어, 부동산 자산 관리자 등이 인기 직무로 나타났으며 정보기술(IT) 분야는 웹개

발자, 모바일 개발자, 데이터 엔지니어가 인기 높은 직무로 나타났다(Robertwalters, 2020).

좋은 일자리(right job)는 우리 삶을 풍요롭게 하고, 보다 나은 미래를 꿈꿀 수 있도록 하는 실체 중 하나이다. 그리고 이 시대의 취업준비생은 ‘좋은 일’을 선택할 권리도 있을 것이다. Yoon(2019)는 양질의 일자리를 찾아 취업률을 향상시키고 유지취업률을 높이기 위한 연구를 진행하였다. 딥러닝의 FE-SONN 모델을 이용하여 취업정보 필터링 과정을 거쳐 최적의 취업처를 매칭시키는 시스템으로 적성에 맞지 않는 중도 퇴사자의 원인 분석을 위한 필터링 시스템을 제안하여 유지취업률을 향상시키기 위한 취업전략을 제안했다. Yoon(2019)는 온라인 취업정보 서비스 시장의 기술적 연구를 진행하였다. 시장의 잠재적 성장 및 이슈 파악을 위해 서비스 사용 트렌드, 주요 서비스 형태, 마케팅 방법 등 산업 트렌드 분석과 SWOT 분석을 통해 취업정보 서비스 시장의 선순환 구조의 전략을 제안하였다(Yoon, 2019).

팬데믹(Pandemic)은 급격한 직무 환경의 변화를 가져왔으며 사무실 개념의 변화, 재택근무의 일상화, 성과 기반의 업무 평가 등 뉴노멀(New normal) 시대로 새로운 업무 방식이 자리 잡고 있다(Egbert, 2020). 본 연구는 기업 비즈니스에 필요한 직무를 빠르게 분석하여 변화하는 조직의 환경에 대응하기 위해 취업준비생에 필요한 스펙을 연구하고자 한다.

2.2 TF-IDF(term frequency-inverse document frequency)

오늘날 기업 내에서는 다양한 비정형 데이터

들이 생성되고 있다. 정보검색이나 문서분류에 가중치를 구하는 알고리즘은 문서 간의 비슷한 정도나 문서 내 어떤 단어가 중요한지의 척도 정도가 필요하다. 단순 단어의 출현 횟수를 요소로 만든 분석을 Bag-of-words(BoW)라 하며 물리적인 단어를 판단하여 분석하다 보니 문장의 뜻을 인지하기 힘들다. 문장에서 단어들이 얼마만큼의 정보를 갖고 있는지 판단하기 위해서 빈번히 등장하는 단어라도 여러 문장에 출현된다면 제한할 필요가 있다. 즉, 중요도가 높은 단어는 벡터를 더 크게 만들고 그렇지 않은 단어는 낮은 벡터로 보정하므로 어떤 카테고리에 속하는 글인지를 판단하는 것이다. 이러한 기법을 TF-IDF(term frequency-inverse document frequency)분석이라 한다(Christian et al., 2016; Qaiser and Ali; 2018; Lee et al., 2019).

TF(Term Frequency)는 특정 단어가 문서 안에서 얼마나 자주 나오는가를 나타내며 단어의 출현 빈도를 나타내는 지표이다. 특정 단어의 출현 횟수를 모든 단어의 출현 횟수로 나눈 값이다. 예를 들어, 어떤 문서가 10,000개의 단어로 이루어져 있고 특정 단어가 120번 나온다면, 단어의 TF는 $120/10000=0.012$ 가 된다. 즉, 식 1)과 같으며 TF의 의미는 어떤 문서에서 특정 단어가 자주 나온다면 그러한 단어는 그 문서에 대한 특징을 갖는 단어가 될 것이다.

반면 IDF(Inverse Document Frequency)는 특정 단어가 전체 문서상에 얼마나 자주 나오지 않는가를 나타내는 벡터이다. 단어의 희소성을 나타내는 지표이다. 문서상의 전체 문장 개수를 특정 단어가 포함된 문장 개수로 나눈 값에 로그를 적용한 값이다. 예를 들면 식2)와 같이 적용한다면 만약 10,000개의 단어로 이루어

져 문서에 문장 개수가 1,000개일 때 120번 특정 단어가 100개의 문장에서 나왔다면, 단어의 IDF는 $\log_{10}(1000/100) = \log_{10}(10) = 1$ 이 된다(Salton and Buckley, 1988).

어떤 문서가 D개의 문장, N개의 단어로 구성된다면 특정 단어 t가 n번 나올 때, 특정 단어 t를 포함한 문장이 d개 있다고 하면 다음과 같다.

$$TF = \frac{n}{N} \quad \text{식1)}$$

$$IDF = -\log_{10} \frac{d}{D} = \log_{10} \frac{D}{d} \quad \text{식2)}$$

$$TF-IDF = TF \cdot IDF = \frac{n}{N} \log_{10} \frac{D}{d} \quad \text{식3)}$$

식3) 은 TF와 IDF를 곱한 것으로 전체 문서 들엔 그 단어를 포함한 문서가 적거나 특정 문서 내에서 단어 빈도가 높으면 TF-IDF 가중치가 높아진다. 따라서 TF-IDF를 이용하여 직무별 스펙에서 흔히 나타나는 단어를 제거하거나, 이슈 스펙 키워드가 얼마나 중요한가를 측정할 때에도 사용된다.

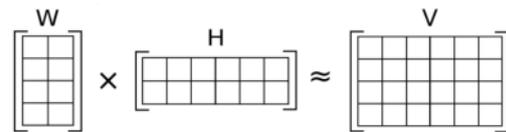
2.3 NMF(Non-negative Matrix Factorization)

기계학습(Machine Learning)은 프로그래밍의 과정 없이 스스로 배우는 능력을 의미한다(Samuel, 1959). 체스를 하는 직업의 성공확률을 체스 게임을 한 경험을 통해서 향상시키는 기술이라고도 볼 수 있다(Jordan and Mitchell, 2015). 기계학습은 크게 데이터 집합을 제공할 때 이 데이터에 정답이 있는 상태를 의미하는 지도학습(supervised learning)과 데이터를 사전

에 분류하지 않고 데이터 셋을 자동으로 분류하는 비지도학습(unsupervised learning)으로 나뉜다. 지도학습은 학습을 시킬 때 정답을 알려주며 학습을 시키며 정답이 있는 문제에 대하여 학습을 시킨다. 주가를 예측, 방문객 예측, 트래픽 수를 예측할 수 있는 기법이며 회귀 분석이 있다. 비지도학습은 정답을 따로 알려주지 않은 상태에서 비슷한 데이터를 군집하여 학습하는 방법으로 뉴스 분류, SNS 관계 분류, 인간 DNA 분류 등이 있다. 가령, AI면접에서도 면접 내용을 보고 합격, 불합격 여부를 나눌 수 있겠지만, 비지도학습을 통해서 면접자를 분류한 뒤 기존 인사팀의 자료를 바탕으로 빠르고 효율적으로 정보를 이용하여 판단할 수 있을 것이다. 주관적인 기준이 들어갈 수 있지만 이런 비지도학습의 특성 때문에 많은 연구자들이 사용하고 있다. 마지막으로 강화학습(reinforcement learning)은 게임에서 일정한 규칙에 따라 행동하여 높은 점수를 얻으면 보상을 주는 것으로 자신이 한 행동에 대한 보상을 받으며 학습을 하는 방법이다. 이는 정해진 결과를 학습하는 지도학습과는 다르며 지향하는 바가 명확하기 때문에 비지도학습이라고도 할 수 없다(Kaelbling, 1996). 강화학습의 대표적인 예로 구글의 인공지능 로봇 알파고(AlphaGo)가 있다.

NMF(Nonnegative Matrix Factorization)는 비지도학습에 속하며 행렬 차원 감소를 이용하기 때문에 비음수 행렬 인수분해라 한다. 문서 토픽의 벡터로 표현하고 LSI(Latent Semantic Indexing)와 Doc2Vec와 비슷하지만 차이가 있다(Kim et al., 2019; Lee, 2020). LSI는 SVD(Singular Vector Decomposition)를 이용하며

각 토픽을 나타내는 새로운 축들이 서로 독립이라 가정하며, 벡터 공간에서 두 벡터가 독립적으로 표현하기 위해서 서로 간의 각도가 90도가 유지되어야 한다. 그리고 한 벡터와 직교인 다른 벡터는 음의 방향 벡터를 가질 가능성이 높다. 이러한 이유로 토픽 벡터의 의미를 해석하기가 어렵다. 이런 문제를 해결하기 위해서 matrix factorization 방법이 NMF이다. Seung and Lee(2001)의 모델인 NMF는 행렬 인수분해 알고리즘을 연구하였다. 원본 행렬 V 를 인수분해 과정을 통하여 두 개의 행렬로 분리하고자 한다. <그림 1>은 원본 행렬 $V = W * H$ 로 분리한 예이다.



<그림 1> Matrix factorization structure in NMF

H 는 특성 행렬로 기존 특성에 대비하여 새로운 특성에 대한 관계를 도출되는 행렬로 나타내며 W 는 가중치 행렬로 특정 카테고리의 특성과의 관계를 활용된다(Berry et al., 2007).

본 연구의 비즈니스 스펙 지수는 TF-IDF 분석을 통해 행렬로 변환하여 특정 상관 계수 이상의 단어를 NMF 알고리즘으로 각 문서의 특성을 추출한다. 추출된 특성을 Danielsson (1980)의 벡터 거리 측정 방법인 코사인 거리 알고리즘을 활용하여 직무별 스펙 유사도를 결정하고자 한다(Danielsson, 1980).

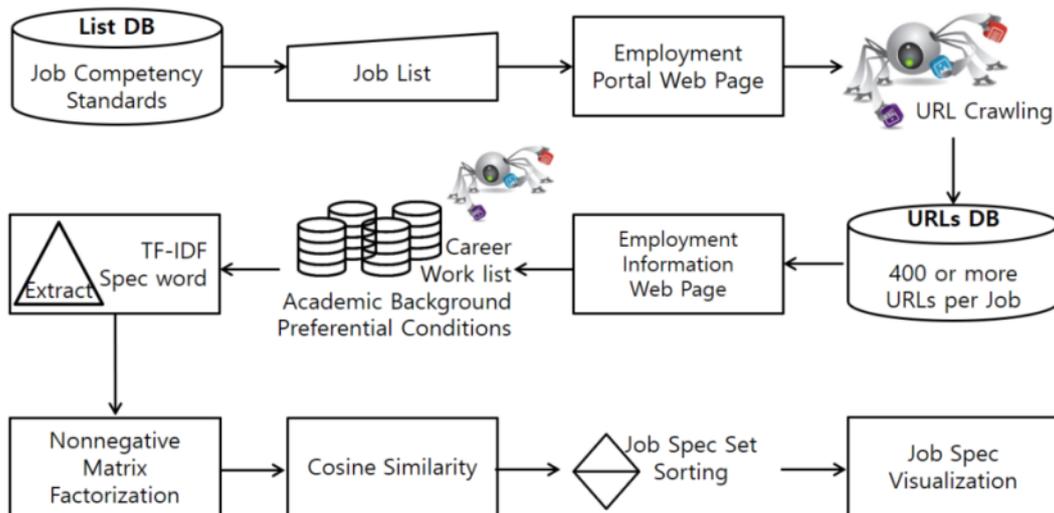
Ⅲ. 연구방법과 프레임워크

정보기술 발달로 인하여 기업 조직, 사람, 문화, 업무 프로세스 변화는 당연한 기업의 생존 전략되었다. 데이터를 기반으로 한 새로운 산업 혁명에서 AI 일자리가 단순노동뿐만 아니라 지식노동, 더 나아가 전문직까지 대체할 가능성과 정보 기술 발달 및 고용시장의 변화로 인한 채용 정보의 부족 때문에 취업 희망자는 이종고를 겪고 있다. 이에 본 연구는 기업의 생존에 필요한 주요한 인재 등용 변화에 대하여 연구하고자 한다. 특히, 데이터 기반 디지털 트랜스포메이션이 급속히 진행되는 현재 상황에서 기업 관점으로 비즈니스에 필요한 직무 리스트를 국가직무능력표준(National Competency Standards) 기반으로 확보할 필요가 있다. 또한 해당 직무가 요구하는 우대조건들에 대한 분석을 통하여 취업 희망자에게 스펙 활동을 지원할 필요가 있다.

직무는 NCS에서 제공하는 한국고용직업분

류(KECO)를 기준으로 대분류, 중분류, 소분류, 세분류(직무)로 구분되며, 직무에 따라 능력단위와 단위요소로 분류되고 있다. 즉, 직무는 능력 단위의 집합이며, 능력단위는 요소, 적용범위, 평가지침, 직업기초능력 등으로 구성되어 있다. 천여 개 이상의 직무 중 경영 교과목 관련 직무를 비즈니스 직무로 선정하였다. 스펙은 Specification의 준말로 하드웨어 제품에 특정 성능과 기능을 나타내기 위한 단어로 쓰이지만 개인이 갖고 있는 능력이나 역량을 표현하는 단어이기도 하다. 본 연구의 스펙은 학력 및 전공을 비롯하여 자격증, 어학 및 연수, 봉사활동, 수상경력, 인턴 등의 개인 역량을 스펙으로 활용하였다.

<그림 2>는 본 연구를 위한 프레임워크이다. 산업 전반에 필요한 직무능력표준에서 일반적인 기업 비즈니스 활동에 필요한 경영, 회계, 사무, 물류 등 관련된 직무를 추출하였다. 추출된 직무 리스트를 채용정보 웹페이지에서 검색하여 직무별 채용정보 링크를 추출하였다. 또한,

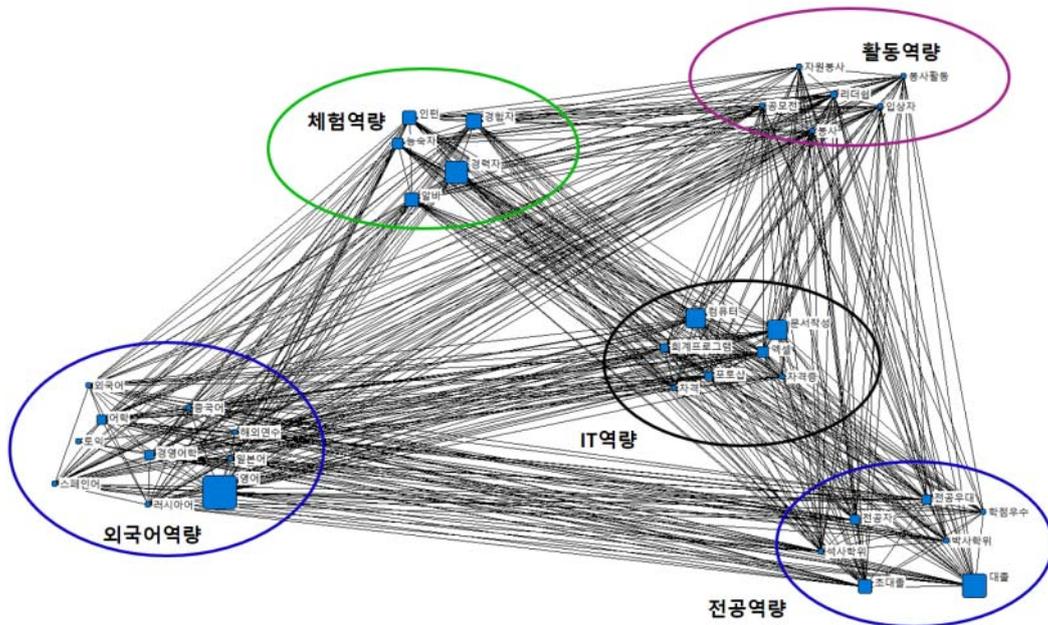


<그림 2> Research framework

직무별 400건 이상의 채용정보 웹페이지 리스트를 DB화하였다. 이후 각각의 채용정보를 접속하여 채용공고 내의 경력, 학력, 우대조건, 자격요건, 업무리스트 등의 데이터를 추출하여 저장하였다. 직무별 추출된 채용정보에 스펙 관련 단어가 채용정보 내에서 얼마나 중요한 것인지를 나타내기 위해 TF-IDF 방식으로 키워드를 추출하였다. 추출된 키워드를 매트릭스로 바꿔서 단어 간 상관관계수 행렬을 통하여 특정 스펙 단어와 연관된 단어의 상관관계수를 이용하여 단어를 선정하였다. 또한 추출된 키워드를 이용하여 채용정보와 직무별 스펙 행렬을 인수분해 과정의 차원 축소 방식(NMF)으로 특성을 추출하였으며, 추출된 벡터 간 거리는 빈도 제한을 고려한 코사인 유사도를 통하여 직무별 스펙의 가중치를 추출하고자 한다.

IV. 실험 및 연구 결과

본 연구는 기업 활동에 필요한 과업 및 작업의 종류와 수준이 비슷한 업무를 모아 비즈니스 직무를 선정하였으며, 직무별 채용정보 분석을 통해 기업에서 필요로 하는 우대 스펙을 발굴하고자 한다. 채용정보에 빈번히 나타난 스펙 정보를 제한하여 의미 있는 스펙 단어를 부각하기 위한 TF-IDF를 사용하였고, 각 채용정보의 특성값을 추출하기 위한 방법으로 NMF의 차원 축소를 진행하였으며, 축소된 차원의 벡터 사이 위치 정보를 각도로 환산한 코사인인자로 적용하여 빈도의 차이에 영향을 최소화하였다. 분석 환경은 Linux CentOS 7.x 서버 환경에서 오픈 소스 소프트웨어인 Linux R를 이용하여 진행되었다. 데이터 수집은 파이썬(python) 셀레니움 웹 크롤링(selenium web crawling)을 이



<그림 3> Competency clustering according to job requirements

<표 1> List of job requirements according to competency

역량	구직요건 속성
전공역량	전공자(0.82), 학점우수(0.72), 대졸(0.67), 초대졸(0.66), 석사학위(0.65), 박사학위(0.63), 전공우대(0.60)
외국어역량	해외연수(0.83), 어학연수(0.75), 영어(0.74, 토익), 경영어학(0.72), 중국어(0.60), 러시아어(0.52), 외국어(0.47), 일본어(0.47), 스페인어(0.45)
IT역량	회계프로그램(0.86), 포토샵(0.85), 컴퓨터(0.76), 엑셀(0.73), 자격(0.64), 자격증(0.62), 문서작성(0.57)
활동역량	자원봉사(0.71), 봉사활동(0.71), 리더쉽(0.6), 봉사(0.53), 입상자(0.51), 공모전(0.51)
체험역량	경험자(0.77), 능숙자(0.74), 경력자(0.72), 인턴(0.67, 인턴경력), 알바(0.60, 아르바이트)

용하여 가상 웨드라이버를 활용하였다.

Jo and Kim(2014)연구와 Kim(2020) 연구에 8대 구직요건(SPEC)을 채용정보와 매칭하여 재구성하였다(Kim, 2020). 특히, Jo and Kim(2014)연구는 학벌, 학점, 어학점수, 해외연수, 자격증, 봉사활동, 기업체험, 수상경력 등의 스펙으로 취업준비생의 인식이 있다고 한다. 또한, 구직요건 속성을 이용하여 CONCOR 분석을 통하여 군집화 과정을 거쳤다. 그 결과 <그림 3>와 같은 역량 간 구직요건 속성들의 군집을 확인할 수 있다. ‘전공역량’, ‘외국어역량’, ‘IT역량’, ‘활동역량’, ‘체험역량’으로 명명하였으며 특히, 활동역량은 봉사, 입상 등의 스펙 활동에 대하여 군집되었으며 체험역량은 경력, 알바 등의 스펙 활동으로 군집되었다.

본 연구는 2019년 1월부터 12월까지 16,000여 개의 채용정보 분석을 통하여 전공역량, 외국어역량, IT역량, 활동역량, 체험역량으로 나누어지는 것을 확인할 수 있었다. <표 1>은 본 연구에 다섯 가지 역량에 구직 요건 속성을 나열한 리스트이다. 전공역량은 ‘전공자’, ‘학점 우수’, ‘대졸’, ‘초대졸’ 등의 순으로 기업 관점의 우선 순위를 확인할 수 있었다. 외국어역량은 ‘어학연수’, ‘토익’, ‘경영어학’, ‘해외연수’,

‘영어’ 등의 키워드가 해당 역량에서 높은 상관관계를 보이고 있다. 또한, ‘회계프로그램’, ‘포토샵’, ‘컴퓨터’, ‘엑셀’, ‘자격증’ 등의 키워드가 IT역량에 해당하였으며, 활동역량은 ‘자원봉사’, ‘봉사활동’, ‘리더쉽’ 등과 높은 관련성을 보였다. 체험역량은 ‘경험자’, ‘능숙자’, ‘경력자’, ‘인턴’, ‘알바’ 등의 키워드를 확인할 수 있었다.

<표 2> List of business jobs according to NCS major classification

대분류	중분류	소분류	직무(스펙 빈도)
사업관리	사업관리	프로젝트 관리	프로젝트관리(30) 산학협력관리(23)
		해외관리	해외법인설립관리(37) 해외취업관리(11)
경영·회계·사무	기획사무	경영기획	경영기획(45) 경영평가(38)
		홍보·광고	PR(21) 광고(35)
		마케팅	마케팅전략기획(45) 고객관리(17) 통계조사(51)
	총무·인사	총무	총무(23) 자산관리(71) 비상기획(64)
	인사·조직	인사(31) 노무관리(55)	
	일반사무	비서(32) 사무행정(37)	

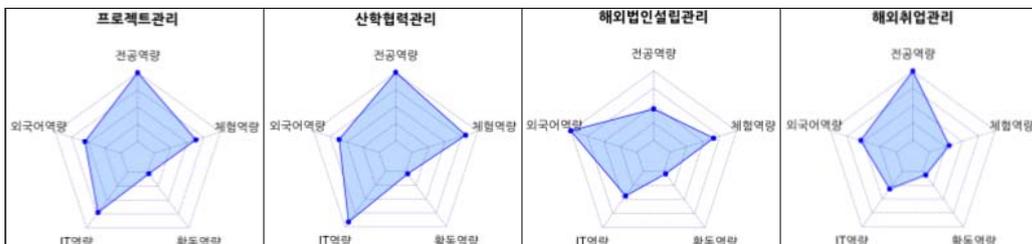
	재무·회계	재무	예산(38) 자금(35)
		회계	회계_감사(39) 세무(24)
	생산·품질관리	생산관리	구매조달(73) 자재관리(37) 공정관리(49) SCM(44)
		품질관리	QM·QC관리(63)
영업판매	영업	일반·해외영업	일반영업(57) 해외영업(51)
		판매	e-비즈니스 전자상거래(41)
	일반판매	매장판매(20) 방문판매(40)	

<표 2>는 산업현장에서 직무에 필요한 능력을 국가가 표준화한 국가직무능력표준(National Competency Standards)의 일부이다. NCS에서는 대분류 24개 분야로 나누어 직무를 정하고 있으며 본 연구는 NCS 대분류 기준 ‘사업관리’, ‘경영회계사무’, ‘영업판매’의 분야에서 기업이 비즈니스에 필요한 직무를 구분하여 나열하였다. 기업 관리에 필요한 비즈니스 관련 직무는 프로젝트관리, 해외관리, 경영기획, 홍보, 광고, 마케팅, 총무, 인사, 노무관리, 일반사무, 재무, 회계, 생산관리, 품질관리, 무역, 유통

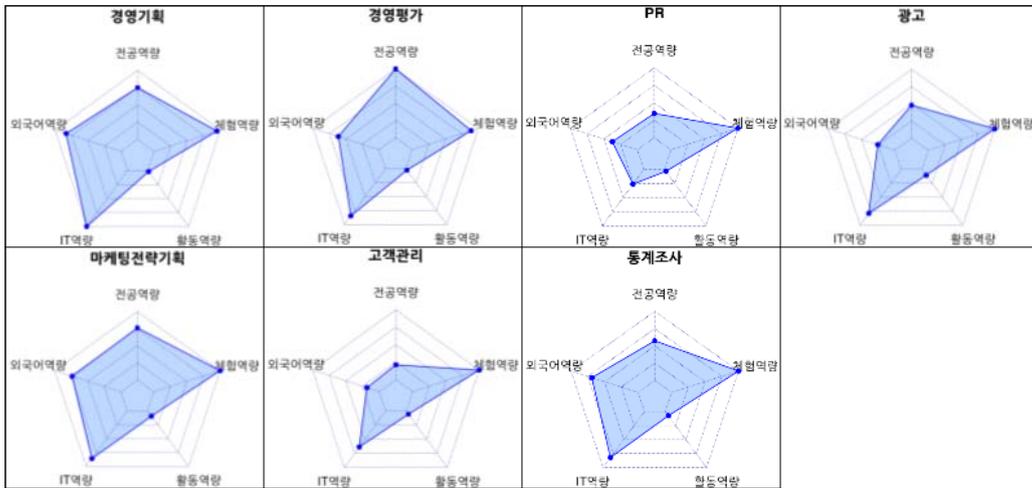
관리, 일반영업, 해외영업, e비즈니스, 일반판매 등 37개로 결정하였다. 본 실험에서는 16,000여개의 채용정보 내 상세내용에서 분석이 이루어졌으며 직무별 스펙 키워드를 추출하여 주요 스펙이 도출되었다. TF-IDF로 추출된 각 직무별 키워드 종류는 직무명과 함께 표현하였으며 직무 ‘PR’인 경우 “능력우수자”, “능숙자”, “마케팅”, “마케팅기획”, “문서작성”, “바이럴마케팅”, “분석”, “브랜드”, “시장분석”, “알바”, “언론홍보”, “온라인마케팅”, “우수자”, “웹마케팅”, “인턴”, “전시”, “컴퓨터활용능력”, “포토샵”, “프리젠테이션”, “홍보”, “홍보기획” 등 21개의 스펙 키워드를 결정하였다. 스펙 빈도는 TF-IDF 방식으로 추출된 키워드를 매트릭스의 단어 간 상관계수 행렬을 통하여 특정 스펙 단어와 연관된 단어의 상관계수를 정렬하여 0.4 이상의 키워드만 선정하였으며 이것이 직무별 스펙 단어 빈도수이다.

<그림 4, 5, 6, 7, 8>는 기업 비즈니스 활동에 필요한 직무의 역량을 시각화하였다. 즉, <표 1>의 구직요건 키워드를 이용하여 <표 2>의 37개 비즈니스 직무 스펙에 대해 1~5점 사이의 방사형 그래프로 각 역량을 시각화하여 정리한 부분이다.

<그림 4>는 사업관리 측면의 4가지 직무 중 프로젝트관리직무는 기업 관점에서 전공역량



<그림 4> Five competency indexes by business job(Business management)



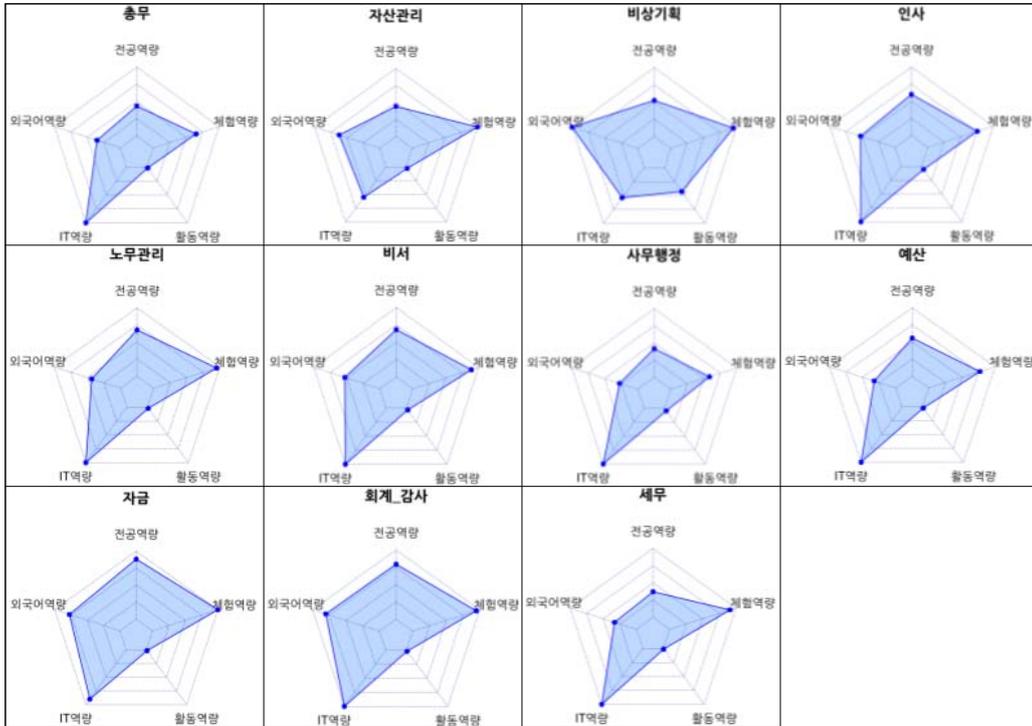
<그림 5> Five competency indexes by business job(Planning affairs)

과 IT역량을 채용정보의 우대조건으로 설정하였으며, 활동역량은 영향력이 낮은 것으로 나타났다. 산학협력관리직무는 IT역량과 전공역량, 체험역량이 높은 것으로 나타났으며 활동역량은 낮았다. 해외법인설립관리직무는 외국어역량과 체험역량이 높은 것으로 나타났으며, 해외취업관리직무는 전공역량만이 높은 것으로 나타났다. 사업관리 직무에서 IT역량과 전공역량이 공통적으로 높게 나타난 것은 여러 사업을 관리하고 프로젝트를 수행하기 때문에 신입 때부터 여러 가지 상황에 잘 대처할 수 있도록 IT역량이 필요하다고 판단된다. 또한 관리자 못지 않은 역할이 필요하기 때문에 업무에 필요한 관련 전공자를 선호하는 것으로 파악된다.

<그림 5>는 기획사무의 7가지 직무 중 경영기획직무, 경영평가직무, 마케팅전략기획직무는 활동역량을 제외하고 모든 역량이 필요한 것으로 나타났다. PR직무, 광고직무, 고객관리직무는 체험역량만이 높게 나타났다. 통계조사직무는 체험역량이 가장 높았으며 IT역량, 외국어역량, 전공역량 순으로 나타났다. 기획사무에

서 활동역량을 제외하고 모든 역량이 필요한 것으로 나타난 것은 기획사무 직무는 조직전체를 파악할 수 있는 밸런스 감각과 사업전략을 바탕으로 기획하고 관리해야하기 때문이다. 또한 기업 전면에 나서지는 않지만 뒤에서 서포트하는 중요한 존재가 기획사무 직무이고, 풍부한 지식과 논리적인 사고가 요구되기 때문에 활동역량을 제외한 모든 역량이 필요한 것으로 판단된다.

<그림 6>은 총무인사와 재무회계의 분석 결과이다. 총무인사는 7가지 직무 중 총무직무, 인사직무, 노무관리직무, 사무행정직무, 비서직무는 IT역량을 가장 높게 기업관점에서 우대조건으로 설정하였으며 자산관리직무는 체험역량, 비상기획직무는 외국어역량, 비서직무는 IT역량과 체험역량이 높은 구직자를 우선적으로 채용하고자 하였다. 재무회계의 4가지 직무 중 예산직무와 세무직무는 IT역량과 체험역량의 영향력이 높았으며 자금직무와 회계_감사직무는 전공, 체험, 외국어, IT 등의 역량을 높게 평가하고 있었다. 총무는 관리부문의 중심이며,

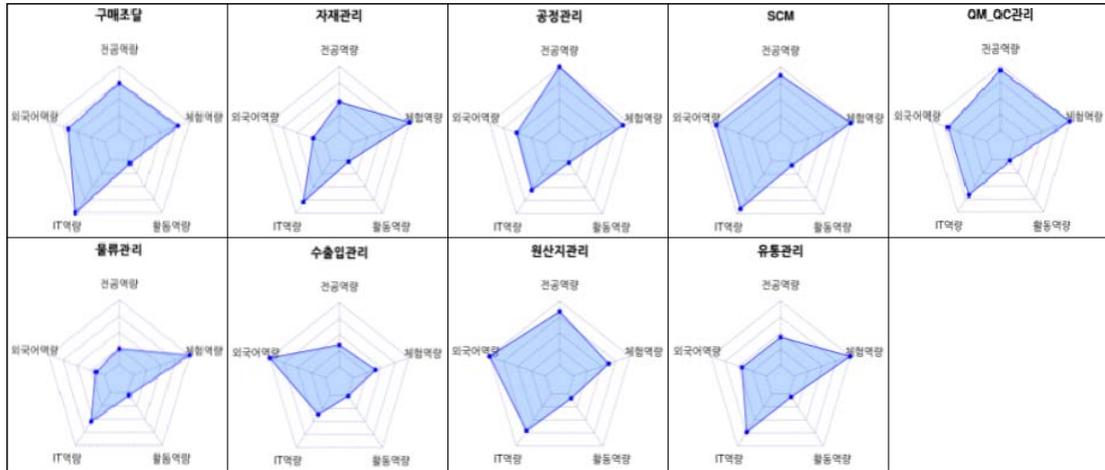


<그림 6> Five competency indexes by business job(HR, Accounting)

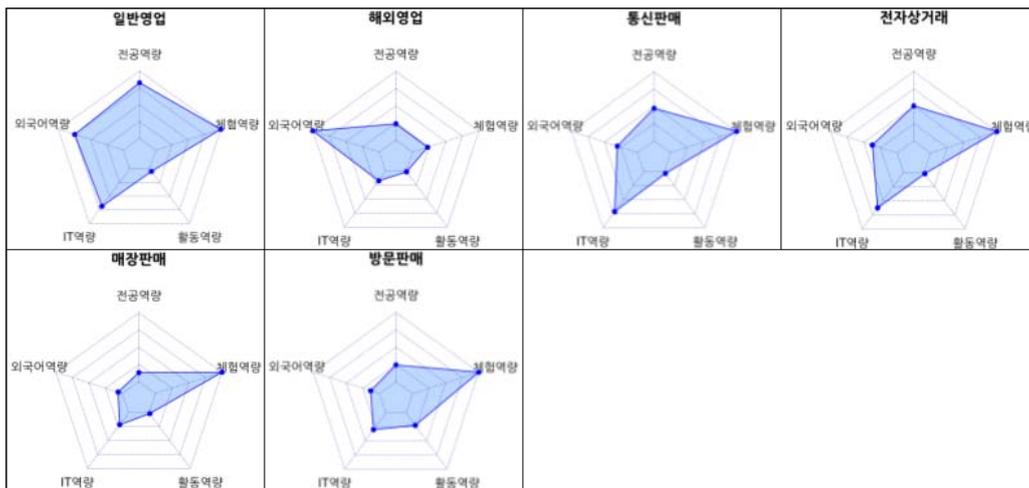
인사는 종업원의 모집이나 채용에 관여하기 때문에 사업전략을 바탕으로 인재를 채용하는 선견성이 필요하다. 또한 재무회계는 기업 자본 자체를 관리하는 직종이기 때문에 사고력과 분석력뿐만 아니라 IT역량에 속하는 표 계산 스킬 등이 이구된다. 따라서 총무인사, 재무회계는 IT역량과 체험역량이 높게 나타난 것으로 판단된다.

<그림 7>은 생산품질관리와 무역유통관리의 분석 결과이다. 생산품질관리의 5가지 직무 중 구매조달직무, 공정관리직무, SCM직무는 활동역량을 제외하고 대부분의 역량이 높은 영향을 미쳤으며 자재관리직무는 체험역량, QM·QC관리직무는 체험과 전공역량을 우대하고 있었다. 무역유통관리의 4가지 직무 중 물류관리직무는

체험역량만이 높았으며 수출입관리직무는 외국어역량만이 높았다. 원산지관리직무는 외국어역량과 전공역량, IT역량순으로 나타났으며 유통관리직무는 체험역량과 IT역량이 높은 것으로 나타났다. 수출입관리직무와 원산지관리 직무를 제외한 생산, 품질관리 업무의 체험역량이 높게 나타난 것은 생산관리 특성상 ‘안전’과 ‘불량을 제로’가 중요하기 때문에 능숙한 경력자를 중심으로 채용을 하게 된다고 볼 수 있다. 또한 제조현장은 효율화를 목적으로 하므로 IT역량 중 품질경영 기사와 같은 품질, 생산관리 직무와 관련된 자격증 취득자나 각 분야의 전공자들을 선호하기 때문에 생산, 품질관리 직무에서는 체험역량과 IT역량, 전공역량이 비교적 높게 나타난 것으로 판단된다.



<그림 7> Five competency indexes by business job(Production, Quality control)



<그림 8> Five competency indexes by business job(Sales)

마지막으로 <그림 8>은 영업판매 6가지 직무 중 일반영업직무는 전공, 체험, IT, 외국어역량이 필요하였으며, 해외영업직무는 외국어역량만이 높았다. 통신판매직무와 전자상거래직무는 체험역량과 IT역량이 높았으며 매장판매직무, 방문판매직무는 체험역량만을 우대하고 있었다. 영업판매는 상품이나 서비스를 소개하고 판매하는 직무이기 때문에 높은 커뮤니케이

션 능력이 필수라고 볼 수 있다. 또한 기업의 이익과 직결되는 중요한 업무이므로 다양한 고객에게 기업의 긍정적 이미지를 보여주기 위하여 풍부한 지식을 바탕으로 고객의 필요에 응답할 수 있는 제안력과 끈질김이 필요하다. 따라서 영업판매에서는 많은 경험이 축적된 경험자를 우대하는 것으로 판단된다.

V. 결론 및 향후 연구과제

혁신적인 정보기술의 수용은 인간과 기술의 결합이라는 4차 산업혁명의 핵심적인 특성에 의하여 더욱 가속화되고 있다. 기술의 응용은 가파른 변곡점을 만들며 기업의 외부환경을 변화시키고 있다. 기업의 변화는 채용시장의 변화, 업무 프로세스, 조직 구조 등의 분야에 연쇄적인 변화를 일으키고 있다. 특히, 기업은 급격한 기술 변화를 수용하고 응용할 수 있는 인재의 수급이 필수적이며, 이에 부합하는 인재들의 확보는 조직의 대응력과 생존에 필수불가결한 요소라고 판단된다. 그러나 과연 기업은 기술 변화와 수용에 부합하는 인재들을 채용하고 있는가 하는 점에서 채용정보의 분석 가치를 확인할 필요가 있었다.

본 연구를 정리하면 다음과 같다. 데이터 전처리 과정에 TF-IDF 알고리즘을 활용함으로써 전체 문서에서 높은 빈도의 단어는 특징에서 배제하고, 문서에 특정 단어가 많이 언급될수록 그 단어의 가중치 값을 크게 하여 문서의 특징을 찾고자 하였다. 추출된 키워드를 매트릭스 상관관계수 행렬을 통하여 특정 스펙 단어와 연관된 단어의 상관계수를 정렬하여 0.4이상의 키워드만 선정하여 불필요한 키워드를 전처리 하였다. 또한, 특성을 해석하기에 불명확한 t-SNE나 PCA의 함의 한계점을 고려하여 본 연구는 NMF 알고리즘을 이용하여 특성 해석이 가능하며 행렬로 압축되어있는 특성을 기반으로 유사도를 측정하였다. 코사인 유사도를 적용하여 직무 스펙 키워드 간 거리에 적용하여 유클리드 거리를 이용한 유사도 측정의 한계점을 보완하였다.

데이터 기반의 디지털 트랜스포메이션이 급속히 진행되고 있는 현재 기업 관점에서 볼 때, 변화하는 비즈니스에 필요한 직무 리스트를 확보하고 해당 직무가 요구하는 역량과 조건들을 적절하게 제시하였다. 또한 산업과 기업의 비즈니스 특성, 직무역량에 대한 분석을 통하여 비즈니스 직무에 필요한 우선 구직요건을 개발하였다. 본 연구를 위하여 비즈니스 직무를 먼저 확정하였다. <표 1>의 역량에 따른 구직요건 속성을 나타낸 결과이다. 역량 구분 없이 40여 개의 속성 전체를 대상으로 상관분석을 해보면 경력자(0.84), 전공자(0.81), 대졸(0.81), 경영어학(0.80), 컴퓨터(0.79), 초대졸(0.79), 알바(0.77), 인턴(0.77), 경험자(0.76)등의 순으로 나타났다. 체험역량과 전공역량의 스펙 속성이 기업 관점에서 요구하는 내용으로 볼 수 있다.

비즈니스 직무는 사업관리, 기획사무, 총무, 인사, 재무, 회계, 생산관리, 품질관리, 영업, 판매 등의 기준으로 분류되었으며 경영기획, 광고, 고객관리, 자산관리, 노무관리, 행정, 자금, 회계, 세무, 구매조달, 공정관리, SCM, 품질관리, 물류, 국내외영업, 전자상거래 등 37개의 직무 리스트를 확정하였다. 37개 리스트 기준에 최소 각 직무별 채용 정보 웹페이지 16,000여 건을 분석하였으며 직무 평균 400여 건으로 집계되었다.

<그림 4, 5, 6, 7, 8>의 연구결과를 종합적으로 살펴보면 다음과 같다. 채용정보는 기업이 인재 등용에 있어서 개인의 역량과 조직의 융합이 조화롭게 이루어질 수 있도록 필요한 인력을 채용하기 위한 정보이다. 기업 활동에 필요한 37개 직무를 분석한 결과 전공역량을 우대하는 직무는 프로젝트관리, 산학협력관리, 해

외취업관리, 경영평가, 공정관리, QM_QC관리 등으로 나타났다. 이는 전공자의 학점우수자를 우선적으로 채용하겠다는 기업의 의지로 판단된다. 체험역량을 우대하는 직무는 경영기획, PR, 광고, 마케팅전략기획, 고객관리, 통계조사, 자산관리, 비상기획, 노무관리, 자금, 회계_감사, 세무, 자재관리, SCM, QM_QC관리, 물류관리, 유통관리, 일반영업, 통신판매, 전자상거래, 매장판매, 방문판매 등으로 나타났다. 이는 경험과 경력, 인턴 등 기업체험을 우선적으로 채용하고자하는 의미로 해석된다. 외국어역량을 우선 고려하여 채용하고자 하는 직무는 해외법인설립관리, 비상기획, SCM, 수출입관리, 원산지관리, 해외영업 등으로 조사되었으며 어학, 영어(토익) 등의 구직요건을 우선으로 고려하는 것으로 나타났다. IT역량은 전체적으로 모든 직무에서 필요로 하는 것으로 나타났으며 특히, 산학협력관리, 경영기획, 광고, 마케팅전략기획, 통계조사, 총무, 인사, 노무관리, 비서, 사무행정, 예산, 자금, 세무, 구매조달, SCM 직무 등이 요구하는 스펙으로 나타났다. 마지막으로 활동역량은 봉사과 공모전 입상, 외부 활동 등의 키워드가 해당 역량을 확인하는 키워드이지만 비즈니스 관련 직무에 반응이 낮으며 비상기획 직무에만 반응을 보이고 있는 특징이 있다.

본 연구는 채용 정보를 분석하여 비즈니스 관련 직무 스펙 연구하였다. 결과를 종합해보면 컴퓨터 활용 및 자격증 관련 스펙인 IT역량과 경력 및 인턴 스펙인 체험역량을 기업이 선호하는 역량이라 판단된다. 또한, 직무에 따라 외국어역량을 추가적으로 선호하는 것으로 확인되었다.

본 연구의 시사점은 다음과 같다. 기업 관점의 스펙 연구는 유사한 산업이나 직무 관점의 채용 우대 정책을 비교함으로써 급격한 기술 변화를 수용하고 응용할 수 있는 인재 수급 기준으로 활용될 것이며, 이에 부합하는 인재들의 확보는 조직의 대응력과 생존에 필수불가결적 요소로 활용될 수 있다. 또한, 취업 희망자 관점의 스펙 연구는 취업을 준비하는 과정에서 스펙 활동 순위 결정은 물론이고 개인 역량 비교를 통하여 폭넓은 직무 지원이 가능할 것으로 본다.

본 연구의 전공역량, 외국어역량, IT역량, 활동역량, 체험역량 등 5대 역량 지수 연구와 구직요건 속성 결정 연구는 빈도 중심 연구를 벗어나 특정 있는 단어에 비중을 두고 연구를 한다면 고객의 니즈 해석에 신뢰도가 높아질 것으로 본다. 하지만 변화하는 고객의 니즈 분석을 통하여 지속적인 스펙 키워드 관리가 필요하다. 특히, 단어를 벡터로 변환하거나 채용정보 내 스펙 단어의 연관성을 표현할 때, 단어 벡터 차원을 축소할 때 즉, 다양한 워드 임베딩(word embedding)기법을 활용하여 순위 결정을 해야 할 것이다. 또한, 지속적인 채용정보 분석은 시계열 데이터의 축적과 중단 연구의 초석으로 직무별 채용 변화 즉, 채용 트렌드를 찾을 수 있을 것이다.

새로운 산업혁명 속에서 정보기술의 변화가 인재 채용에 얼마나 영향을 끼치는지에 대한 지속적 연구도 가능할 것으로 본다. 인공지능의 발달과 RPA(Robotic Process Automation) 산업의 성장으로 취업 희망자의 이중고는 기술의 변화가 지속되는 동안 계속될 것으로 판단되며 채용정보 연구는 전 산업에 걸쳐 확대되어야

할 것이다. 또한, 기업 채용 담당자의 인터뷰를 통하여 채용정보에 나타나지 않은 인성 역량 분석 연구도 필요할 것으로 사료된다.

참고문헌

- Berry MW, Browne M, Langville AN, Pauca VP, and Plemmons RJ., “Algorithms and Applications for Approximate Nonnegative Matrix Factorization,” *Computational statistics data analysis*. Vol. 52, No. 1, 2007, pp 155-173.
- Castells, Manuel., “The Rise of the Network Society,” New York: Wiley, 2010.
- Christian, H., Agus, M. P., and Suhartono, D. “Single Document Automatic Text Summarization Using Term Frequency-Inverse Document Frequency (TF-IDF),” *ComTech: Computer, Mathematics and Engineering Applications*, Vol. 7, No. 4, 2016, pp. 285-294.
- Danielsson, P. E., “Euclidean Distance Mapping,” *Computer Graphics and image processing*, Vol. 14, No. 3, 1980, pp. 227-248.
- Egbert J., “The New Normal?: A Pandemic of Task Engagement in Language Learning,” *Foreign language annals*. Vol. 53, No. 2, 2020, pp. 314-319.
- Graham, M., Lehdonvirta, V., Wood, A., Barnard, H., Hjorth, I., ' and D Simon, P., “The Risks and Rewards of Online Gig Work at the Global Margins”, 2017.
- Heeks, R., “Decent Work and the Digital Gig Economy: a Developing Country Perspective on Employment Impacts and Standards in Online Outsourcing, Crowdwork, Etc.,” *Development Informatics Working Paper*, Vol. 71, 2017.
- ILO. Available online : https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_771909/lang--en/index.htm (accessed on 23 February 2021).
- ilo.org. Available online : https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_762547/lang--en/index.htm (accessed on 02 December 2020).
- Jordan, M. I., ' Mitchell, T. M. “Machine Learning: Trends Perspectives, and Prospects,” *Science*, Vol. 349, No. 6245, 2015, pp. 255-260.
- Kaelbling, L. P., Littman, M. L., and Moore, A. W., “Reinforcement Learning: A Survey,” *Journal of Artificial Intelligence Research*. Vol. 4, 1996, pp. 237-285.
- Kim D, Seo D, Cho S, Kang P., “Multi-co-training for Document Classification Using Various Document Representations: TF-IDF, LDA, and Doc2Vec,” *Information Sciences*, 2019, pp.15-29.
- Kim Y. H., “A Study on the Perception of

- Recruiters Selected through NCS Based Recruitment System: Newly Hired Employees in 25 Public Organization,” *Journal of Skills and Qualifications*, Vol. 9, No. 2, 2020, pp. 17-40.
- Lee J. H., “Research on Designing Korean Emotional Dictionary Using Intelligent Natural Language Crawling System in SNS,” *The Journal of Information Systems*, Vol. 29, No. 3, 2020, pp. 237-251.
- Lee J. H., Lee M. B., and Kim J. W., “A Study on Korean Language Processing Using TF-IDF,” *The Journal of Information Systems*, Vol. 28, No. 3, 2019, pp. 105-121.
- Kaiser, S., and Ali, R., “Text Mining: use of TF-IDF to Examine the Relevance of Words to Documents,” *International Journal of Computer Applications*. Vol. 181, No. 1, 2018, pp. 25-29.
- Robertwalters. Available online : <https://www.robertwalters.co.kr/hiring/hiring-trends-2020.html>(accessed on 2020).
- Salton, G., and Buckley, C., “Term-weighting Approaches in Automatic Text retrieval,” *Information Processing Management*, Vol. 24, No. 5, 1988, pp. 513-523.
- Samuel AL. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*. Vol. 3, No. 3, pp. 210-229.
- Surie, A., ' Sharma, L. V., “Climate Change, Agrarian Distress, and the Role of Digital Labour Markets: Evidence from Bengaluru, Karnataka,” *Decision*, Vol. 46, No. 2, 2019, pp. 127-138.
- Yoon S. H., “A Study on Employment Strategy Based on Employment Information Filtering,” *The Journal of the Convergence on Culture Technology*. Vol. 5, No. 4, 2019, pp. 251-258.
- Yoon, J. W. “A Descriptive Study on the Job Information Service Market,” *Journal of The Korea Society of Computer and Information*. Vol. 24, No. 11, 2019, pp. 143-152

이 종 화 (Lee, Jong Hwa)



부경대학교 경영학 박사학위를 취득하고, 현재 동의대학교 정보경영학부 e비즈니스학 전공 교수로 재직 중이다. 주요 관심분야는 BigData, Text Mining, Content Analysis, Sentiment Analysis 등이다.

이 현 규 (Lee, Hyun Kyu)



연세대학교에서 경영학 박사학위를 취득하고, 현재 부경대학교 경영학부 교수로 재직하고 있으며, 주요 관심분야는 정보시스템전략, Data-Mining ' Analysis 등이다.

<Abstract>

Research on Business Job Specification through Employment Information Analysis*

Lee, Jong Hwa · Lee, Hyun Kyu

Purpose

This research aims to study the changes in recruitment needed for the growth and survival of companies in the rapidly changing industry. In particular, we built a real company's worklist accounting for the rapidly advancing data-driven digital transformation, and presented the capabilities and conditions required for work.

Design/methodology/approach

we selected 37 jobs based on NCS to develop the employment search requirements by analyzing the business characteristics and work capabilities of the industry and company. The business specification indicators were converted into a matrix through the TF-IDF process, and the NMF algorithm is used to extract the features of each document. Also, the cosine distance measurement method is utilized to determine the similarity of the job specification conditions.

Findings

Companies tended to prefer "IT competency," which is a specification related to computer use and certification, and "experience competency," which is a specification for experience and internship. In addition, 'foreign language competency' was additionally preferred depending on the job. This analysis and development of job requirements would not only help companies to find the talents but also be useful for the jobseekers to easily decide the priority of their specification activities.

Keyword: Text Mining, TF-IDF, Non-negative Matrix Factorization, Specification, Job Competency

* 이 논문은 2021년 11월 24일 접수, 2021년 12월 16일 1차 심사, 2022년 2월 24일 2차 심사, 2022년 3월 29일 게재 확정되었습니다.

* This work was supported by a Research Grant of Pukyong National University(2021).