

구인구직사이트의 구인정보 기반 지능형 직무분류체계의 구축

이정승
호서대학교 경영학부
(leebahk@gmail.com)

주요 구인구직사이트의 직무분류체계가 사이트마다 상이하고 SW분야에서 제안한 ‘SQF(Sectoral Qualifications Framework)’의 직무분류체계와도 달라 SW산업에서 SW기업, SW구직자, 구인구직사이트가 모두 납득할 수 있는 새로운 직무분류체계가 필요하다. 본 연구의 목적은 주요 구인구직사이트의 구인정보와 ‘NCS(National Competency Standards)’에 기반을 둔 SQF를 분석하여 시장 수요를 반영한 표준 직무분류체계를 구축하는 것이다.

이를 위해 주요 구인구직사이트의 직종 간 연관분석과 SQF와 직종 간 연관분석을 실시하여 직종 간 연관규칙을 도출하고자 한다. 이 연관규칙을 이용하여 주요 구인구직사이트의 직무분류체계를 맵핑하고 SQF와 직무분류체계를 맵핑함으로써 데이터 기반의 지능형 직무분류체계를 제안하였다.

연구 결과 국내 주요 구인구직사이트인 ‘워크넷,’ ‘잡코리아,’ ‘사람인’에서 3만여 건의 구인정보를 open API를 이용하여 XML 형태로 수집하여 데이터베이스에 저장했다. 이 중 복수의 구인구직사이트에 동시 게시된 구인정보 900여 건을 필터링한 후 빈발 패턴 마이닝(frequent pattern mining)인 Apriori 알고리즘을 적용하여 800여 개의 연관규칙을 도출하였다. 800여 개의 연관규칙을 바탕으로 워크넷, 잡코리아, 사람인의 직무분류체계와 SQF의 직무분류체계를 맵핑하여 1~4차로 분류하되 분류의 단계가 유연한 표준 직무분류체계를 새롭게 구축했다.

본 연구는 일부 전문가의 직관이 아닌 직종 간 연관분석을 통해 데이터를 기반으로 직종 간 맵핑을 시도함으로써 시장 수요를 반영하는 새로운 직무분류체계를 제안했다는 데 의의가 있다. 다만 본 연구는 데이터 수집 시점이 일시적이기 때문에 시간의 흐름에 따라 변화하는 시장의 수요를 충분히 반영하지 못하는 한계가 있다. 계절적 요인과 주요 공채 시기 등 시간에 따라 시장의 요구하는 변해갈 것이기에 더욱 정확한 매칭을 얻기 위해서는 지속적인 데이터 모니터링과 반복적인 실험이 필요하다. 본 연구 결과는 향후 SW산업 분야에서 SQF의 개선방향을 제시하는데 활용될 수 있고, SW산업 분야에서 성공을 경험삼아 타 산업으로 확장 이전될 수 있을 것으로 기대한다.

주제어 : 연관규칙, 빈발 패턴 마이닝, 직무분류체계, SW산업

논문접수일 : 2019년 11월 13일 논문수정일 : 2019년 12월 22일 게재확정일 : 2019년 12월 23일
원고유형 : 일반논문 교신저자 : 이정승

1. 서론

SW산업은 지속적으로 성장 중이며 미래에 중요한 산업이다. 스마트폰 등장 이후 SW는 개인·기업·국가의 경쟁력을 결정하는 바로미터

로서 작용하는 핵심 산업으로 급부상 중이다. 그런데 SW기업은 구인란을 겪고 있으며, 반대로 SW구직자는 취업난을 겪는 상황이다. SW 개발 전문가의 구인 인원 중 1/3의 충원도 어렵다고 호소한다(Na, 2013).

이러한 ‘구인구직자간 미스매치’는 SW 인력 확보를 어렵게 한다. 이를 해결하기 위해 ‘NCS (National Competency Standards)’를 도입하였으나 일부 공기업을 제외한 일반기업에서는 거의 사용하지 않는다. 더구나 주요 구인구직사이트인 워크넷(WORKNET), 사람인(saramin), 잡코리아(JOB KOREA)의 직무분류체계가 상이하고 SW분야에서 제안한 ‘SQF(Sectoral Qualifications Framework)’의 직무분류체계와도 다르다(Park et al., 2016). 따라서 SW산업에서 SW기업, SW구직자, 구인구직사이트와 같은 시장 참여자가 모두 납득할 수 있는 새로운 ‘직무분류체계’가 필요하다.

본 연구의 목적은 주요 구인구직사이트와 SQF를 분석하여 SW기업, SW구직자, 주요 구인구직사이트를 위해 시장수요 중심의 표준 직무분류체계를 제안하는 것이다. 여기에서 시장수요 중심은 위원회나 몇몇 전문가의 의견이 아닌 주요 구인구직사이트의 최신 구인정보 데이터를 즉시 반영하겠다는 의미이다. 연구 범위는 SW산업 관련 직무에 한정하여 주요 구인구직사이트의 SW 관련 직무분류체계와 SQF를 대상으로 한다.

2. 관련 연구

2.1 SW산업 직무분류체계

직무분류체계에 대한 연구는 주로 일반 연구자가 아닌 정부와 기관을 중심으로 이루어지는데 표준화된 직무분류체계를 개발하기 위해서 많은 시간과 예산이 필요하기 때문이다. 직무분류체계를 구축하는 방법은 거의 대부분 해당 분야의 전문가로 구성된 위원회를 중심으로 협의

를 통해 이루어지는 것으로 파악된다.

SW분야서는 최근 국내외에서 표준 직무분류체계를 개발하기 위한 연구가 활발하게 진행되고 있다. SW산업에 적합한 표준 직무분류체계를 구축하기 위해 한국소프트웨어산업협회는 2016년 SW분야 산업별 역량체계(SQF)를 제시했다(Park et al., 2016). SQF는 국가역량체계(NQF) 구축의 일환으로 SW산업에서 역량체계를 개편할 때 원칙, 방향, 쟁점을 고려하고 NCS를 참고하여 개발되었다. SQF는 SW분야의 직종을 NCS 분류체계를 기준으로 8개의 직종과 22종의 세분류로 설정하였다. 특히 정보보호 분야에서는 정보보호 직무능력을 표준화하고 직무수행에 필요한 지식과 기술에 대한 정의를 통해 실무 인력의 양성에 도움이 되고자 한국정보화진흥원과 협력하여 정보보호 분야 직무체계를 개발하였다(Jun et al. 2009).

해외 동향을 살펴보면, 미국의 NWCET은 직무체계를 8개의 직무군으로 구분하였고, 영국의 SFIA은 정보통신 직무체계를 6개의 카테고리과 17개의 하위 카테고리의 두 단계로 구분해 개발되었다(NWCET, 2003; SFIA, 2005). 또한 일본의 IPA는 11가지의 직무로 구분하여 IT관련 서비스 제공에 필요한 역량을 체계화했다.

2.2 데이터 기반 직무분류체계

기존의 전문가와 위원회 중심의 직무분류체계 개발과 달리 최근 시장의 수요를 반영할 수 있는 데이터와 머신러닝 기법을 이용한 직무분류체계에 대한 연구가 일부에서 진행되고 있다. 전자문서를 지도학습의 대표적인 기법인 ‘나이브 베이즈 모델링(Naive Bayes Modeling)’으로 학습하여 직무를 자동으로 분석했다(Kang and Ko, 2019).

실험결과 전년도에 수작업으로 라벨링한 결과와 비교해 정확도가 81%에 이르러 고무적이었다. 또한 데이터를 활용한 지능형 직무분류체계는 전문가와 위원회가 매년 직무분류체계를 갱신해야 하는 노력과 비용을 줄여주고 시장의 변화에 신속히 반응할 수 있는 장점이 있다.

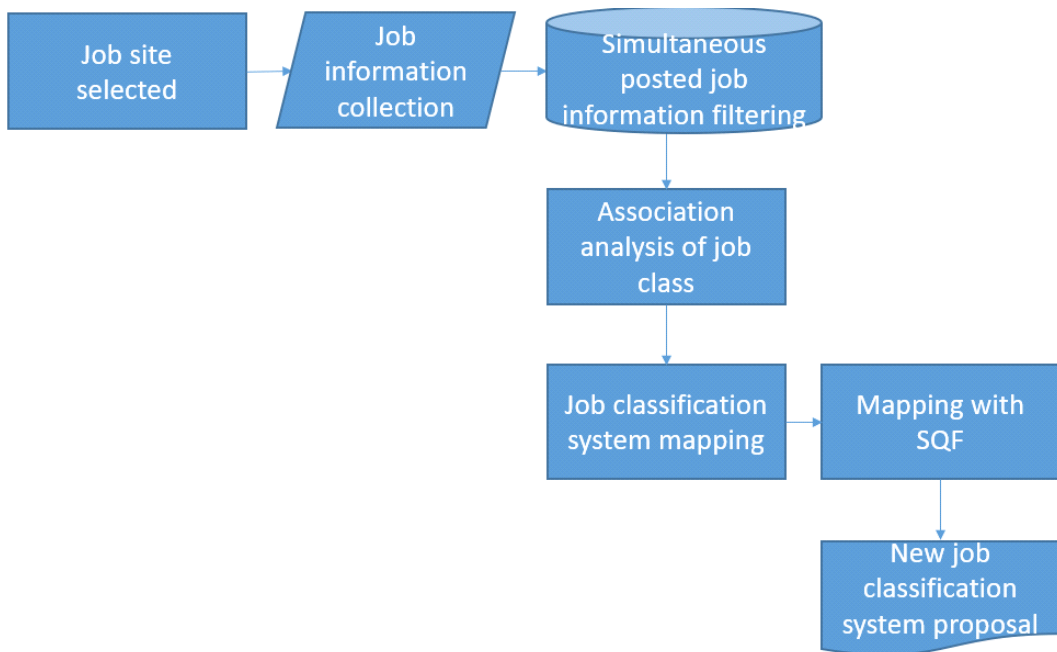
정보통신 분야는 급속한 기술의 변화로 다른 분야에 비해 직업과 직무의 변화가 심하다. 이에 정보관리와 정보서비스 분야의 직업을 대상으로 직무를 분석하고, 직무 간 관련도를 살피고, 직무 중복도를 검토하여 직업 간 이동 가능성과 직업 변화의 동향을 제시했다. KJ 기법 분석결과 5개의 대분류로 나누고, 분류 간 관련 업무를 분석했다(Ahn, 2005).

3. 연구방법론

3.1 연구 추진절차

본 연구는 시장수요 중심의 표준화된 직무분류체계를 제안하는데 목적이 있다. 이를 위해 주요 구인구직사이트의 직종 간 연관분석과 SQF와 직종 간 연관분석을 통해 연관규칙을 찾아낸 후, 주요 구인구직사이트 간 직무분류체계를 맵핑하고 SQF와 직무분류체계를 맵핑할 것이다. 연관규칙을 이용한 맵핑 결과를 바탕으로 시장 중심의 새로운 지능형 직류체계를 제안하고자 한다. 연구 추진절차를 도식화하면 다음 <Figure 1>와 같다.

먼저 SW시장의 직무분류체계에 대한 정보를 얻기 위해 구인구직사이트를 선정한 후, 각 사이



<Figure 1> Research Procedure

트에서 구인정보를 수집하기 위한 방법을 파악하고 Open API를 통해 데이터를 수집할 것이다. 데이터 간 연결 관계에 기반을 두고 각 구인구직 사이트에 동시에 게시된 구인정보만을 남기고 이외의 구인정보는 삭제한다. 이후 직종 간 연관 분석을 수행하고 그 결과물인 연관규칙을 이용하여 구인구직사이트 간의 직무분류체계를 매핑할 것이다. 시장 분류 간의 매핑을 수행한 후, 추가적으로 SQF를 매핑하기 위해 전문가들이 종합적으로 논의하여 최종적으로 새로운 직무분류체계를 제안하고자 한다.

3.2 지능형 직무분류체계 분석

기존의 직무분류체계는 전문가의 의견을 바탕으로 수립하였기에 직무를 분류하는 기준으로서의 가치와 의미가 있다. 그러나 실무적 관점에서 봤을 때 구직자, 구인회사, 중계업자 등 시장에서 사용되고 있는 직무분류와의 괴리가 있을 수 있다. 또한 시간의 흐름에 따라 변화하는 시장의 속도를 맞추기 위해서는 유연한 직무분류체계 조정 사이클이 필요하지만 구직자, 구인회사, 중계업자의 관점이 모두 다르기 때문에 이것을 모두 고려하여 직무분류체계를 만드는 것은 어려운 작업이다.

본 연구의 특징은 데이터 기반 분석으로 볼 수 있는데 동시에 구인구직 사이트에 올라온 구인정보를 기준으로 각각의 분류 체계를 연결하고 하나의 통합된 뷰(view)를 생성할 수 있다는 아이디어는 기존의 직무분류체계 정의를 위한 방법론과는 차별화되어 있다. 실제 직무분류가 활용되는 시장의 관점을 고려할 수 있다.

기존의 직무분류체계와 차별화된 본 연구의 특징은 다음과 같다. 첫째, 구인구직사이트 선정

이다. 공공과 민간부문의 구인구직사이트 중 가장 많이 쓰이고, 포괄적이어서 시장 리더인 구인구직사이트를 주요 분석 대상으로 선정한다. 둘째, 구인정보 수집 방법이다. 각 구인구직사이트에서 제공하는 API를 이용해 JAVA 프로그래밍을 통해 XML형태로 제공되는 구인정보를 수집해 DB로 만들어 저장한다. 셋째, 동시에 게시된 구인정보 필터링이다. MySQL을 이용해 DB에 저장된 구인정보 중 두개 이상의 구인구직사이트에 동시에 게시된 구인정보를 필터링한다. 넷째, 직종 간 연관분석이다. 연관분석을 이용해 동시에 게시된 구인정보가 속한 직종을 연결함으로써 직종 간 연관규칙을 도출한다. MySQL을 이용해 연관규칙의 조건과 결과에 있는 여러 직종의 조합을 해석해 의미 있는 연관규칙을 필터링할 것이다. 다섯째, 직무분류체계 맵핑이다. 직종 간 연관분석으로 도출한 연관규칙을 이용해 각 구인구직사이트의 직무분류체계를 맵핑한다. 여섯째, SQF와 맵핑이다. SQF 기반 구인정보를 필터링하여 이 구인정보가 속한 구인구직사이트의 직무분류체계를 SQF 직무분류체계와 수작업으로 맵핑할 것이다. 일곱째, 새로운 직무분류체계 제안이다. 구인구직사이트의 직무분류체계 매핑 결과와 SQF 직무분류체계 맵핑 결과를 토대로 시장수요 중심의 새로운 표준 직무분류체계 제안하고자 한다.

3.3 분석 대상 및 범위 선정

분석 대상으로 채용공고 정보를 제공하는 주체인 워크넷, 잡코리아, 사람인을 선정하였다. 이 중 워크넷(<http://www.work.go.kr>)은 한국고용정보원에서 운영하고 있으며, 잡코리아(<http://www.jobkorea.co.kr>)와 사람인(<http://www.saramin.co.kr>)

은 사기업으로서 각각 채용정보를 직종별로 분류하여 제공하고 있다. 워크넷은 나라일터, 알리오 등 공공 구인구직 사이트 중 가장 많이 사용되는 대표적인 사이트이기 때문에 선정하였고, 잡코리아와 사람인은 민간 부문에서 시장 리더이기 때문에 선정하였다. 채용정보 사이트에서 자체적으로 제공하는 Open API를 통해 자료를 수집하였는데 연구자의 요청을 통해 API에 접근할 수 있는 키를 획득하고, 자동적으로 데이터를 수집할 수 있도록 자바 프로그램을 구현하였다. 각 사이트에서는 여러 직종에 대한 채용정보를 가지고 있으나, 본 연구의 목적에 맞게 분석 범위를 SW 관련 직군으로 한정하였다. 이를 위해 워크넷은 ‘IT·정보통신·웹 직종’으로, 잡코리아는 ‘IT·인터넷 직종’으로, 사람인은 ‘IT·인터넷 직종’으로 설정하였다.

4. 연구 결과

4.1 구인정보 데이터 수집

워크넷 서비스는 오픈 API를 통해 채용정보, 직업정보, 학과정보를 제공하고 있다. 본 연구에서 활용한 API는 채용정보 API로, 워크넷이 보유하고 있는 일자리 정보(지역별, 직종별 등 조건에 맞는 일자리 정보)를 인증키를 신청하여 발급받은 각 기관(개인 제외)에서 활용할 수 있도록 했다.

잡코리아의 오픈 API는 채용공고 2시간 주기로 업데이트되며, 업·직종, 지역, 학력 등 세부 조건을 설정하여 최대 5,000개의 채용공고 획득 가능하다. 주요 요청 파라미터는 ‘size(출력 건수),’ ‘page/start(페이지 정보),’ ‘rbc(업·직종 대

분류),’ ‘rpcd(업·직종 소분류)’와 같으며, 업·직종 대분류를 IT·인터넷 직종으로 설정(rbcd = 38888)하고 소분류는 해당 대분류의 하위에 위치한 11개 소분류로 한정하여 검색하였다. 즉 대분류를 설정한 후 소분류를 반복적으로 투입하면서 데이터를 수집하였다.

사람인은 개발자가 채용정보 데이터베이스를 사용하여 맞춤형 채용정보 검색 사이트, 채용정보 위젯, 매쉬업 및 채용안내 관련 인터넷 어플리케이션을 개발할 수 있도록 취업 검색 API를 제공하고 있다. 특징적인 점으로는 현재 베타 서비스로서, 인증키를 발급받을 필요가 없이 사이트에서 기본적으로 제공하는 URL을 활용하여 누구나 접근할 수 있다. 취업 검색 API를 활용하여 데이터를 수집하고 저장하였는데 요청 파라미터 중 직종 코드를 IT·인터넷 직종(job_category=4)으로 한정하여 요청하였다.

API 요청에 대한 반환 결과는 XML 형식으로 제공되며 출력 필드의 태그 구조는 <Figure 2>와 같다.

2016년 3월 16일, 10:46 기준 데이터 수집 건수는 다음 <Table 1>과 같다. API를 이용해 구인정보를 수집 시, 웹사이트에서 보이는 것과 차이가 있음을 확인하였다. 특히 그 차이는 워크넷이 가장 컸으며, API로는 웹사이트에서 제공하는 건수의 11% 가량 밖에 제공하지 못하였다. 개수의 차이는 있으나 내용 확인 결과 API가 잘못 끌고 오는 것은 아니었으며, 단순히 웹사이트에서 제공하는 전체를 다 끌어오지 못하는 것임을 파악하였다. 추가적으로 워크넷의 채용정보검색 기능을 이용해 확인한 결과, 채용정보 검색에서 정보 제공처를 워크넷으로 한정했을 경우의 데이터가 API로 얻은 데이터와 거의 일치하는 것을 확인하였다. 워크넷의 채용정보검색 창에서

<pre> <wanted> <wantedAuthNo></wantedAuthNo> <company></company> <title></title> <salTpNm></salTpNm> <sal></sal> <region></region> <holidayTpNm></holidayTpNm> <minEdubg></minEdubg> <career></career> <regDt></regDt> <closeDt></closeDt> <infoSvc></infoSvc> <wantedInfoUrl/> <zipCd></zipCd> <strtnmCd></strtnmCd> <basicAddr></basicAddr> <detailAddr></detailAddr> </wanted> </pre>	<pre> <Items> <GI_No></GI_No> <C_URL></C_URL> <C_Name></C_Name> <GI_Subject></GI_Subject> <GI_Part_No></GI_Part_No> <GI_Career></GI_Career> <GI_Career_Year_Cnt></GI_Career_Year_Cnt> <GI_Pay></GI_Pay> <GI_Pay_Term></GI_Pay_Term> <GI_Pay_Flag></GI_Pay_Flag> <GI_EDU_CutLine></GI_EDU_CutLine> <GI_Edu_Options></GI_Edu_Options> <GI_Keyword></GI_Keyword> <GI_Pass_Type></GI_Pass_Type> <GI_Job_Type></GI_Job_Type> <Staff></Staff> <Jikgub/> <AreaCode></AreaCode> <GI_End_Date></GI_End_Date> <JK_URL></JK_URL> <JK_QR></JK_QR> <GI_W_Date></GI_W_Date> <GI_E_Date></GI_E_Date> </Items> </pre>	<pre> <job> <id></id> <url></url> <active></active> <posting-timestamp></posting-timestamp> <modification-timestamp></modification-timestamp> <opening-timestamp></opening-timestamp> <expiration-timestamp></expiration-timestamp> <close-type code=""></close-type> <company> <name href=""></name> </company> <position> <title></title> <location code=""></location> <job-type code=""></job-type> <industry code=""></industry> <job-category code=""></job-category> <open-quantity></open-quantity> <experience-level code="" min="" max=""></experience-level> <required-education-level code=""></required-education-level> </position> <keyword></keyword> <salary code=""></salary> </job> </pre>
--	---	---

〈Figure 2〉 Tag Structure of Output Fields of WORKNET, JOBKOREA and saramin

조건을 설정하는 부분을 살펴보면, 가장 아랫부분의 정보 제공처 조건을 워크넷으로 한정하여 검색한 데이터가 API에서 제공하는 데이터임을 알 수 있다.

람인 코드 체계를 그대로 가져오기 때문에 하나의 채용공고에 대한 서로 다른 분류체계를 분석하고자 하는 본 연구에서는 필요하지 않았다.

〈Table 1〉 The Number of Data Collection

No	Appeared on the web	Acquired and saved by API
WORKNET	22,333	2,477
JOBKOREA	> 18,000	10,799
saramin	> 17,000	17,874

초기에는 이 이슈에 의해 워크넷 데이터의 대부분이 사라지게 되므로 문제라고 판단하고 해결을 시도하였으나, 실제로 웹사이트의 채용공고 전체를 끌어오더라도 의미가 없는 것을 확인하였다. 외부 정보 제공처에서 가져온 채용공고의 경우, 채용공고를 제공한 곳의 잡코리아, 사

4.2 데이터 매칭

동일한 구인정보가 서로 다른 분류 체계에서 어떻게 분류되었는지 파악하기 위해 워크넷, 잡코리아, 사람인 데이터를 데이터베이스에 저장하였다. 데이터베이스의 질의문인 SQL을 통해 데이터를 매칭하는 것이 상대적으로 용이하기 때문에 각각 worknet, jobkorea, saramin 테이블에 저장했다.

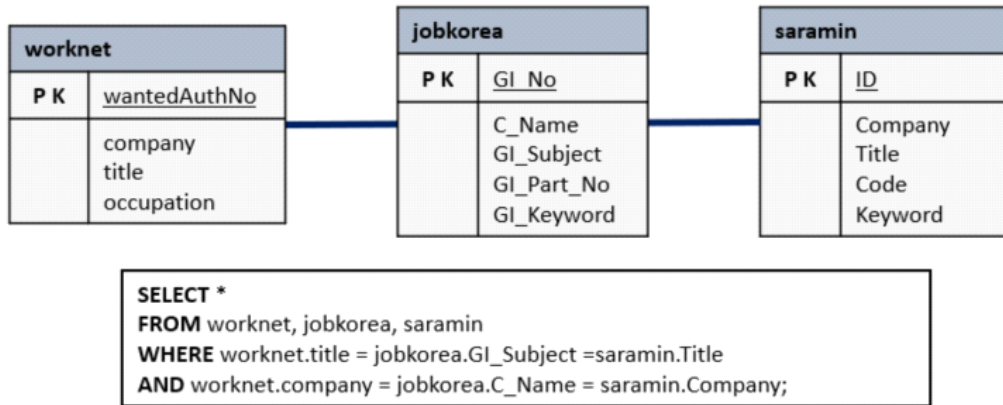
워크넷, 잡코리아, 사람인 동시 게시 구인정보를 필터링하기 위해 데이터베이스 연결 형태와 사용한 쿼리는 다음 〈Figure 3〉과 같다.

여러 데이터의 조합 중, 채용공고명과 회사명이 동일하다면 동일한 채용공고라고 판단할 수 있다. 회사명만으로 동일한 채용공고를 판단할

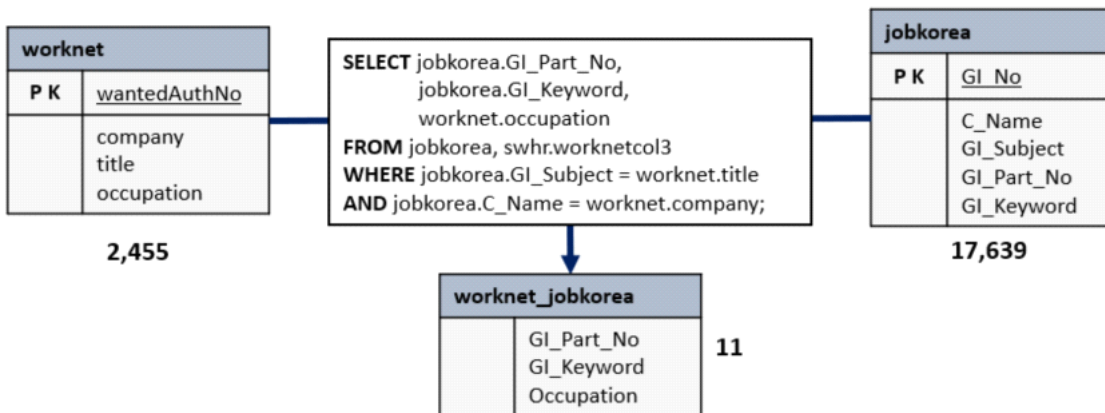
경우, 이름은 같으나 다른 회사의 채용공고가 검색되는 경우가 있었다. 그 외에도 하나의 회사가 채용공고명만 다른 여러 채용공고를 올린 경우도 있기 때문에 회사명만으로는 정확한 매칭이 불가능하다. 필터링 결과 3개 웹사이트에 동시에 게시된 구인정보의 레코드 수는 0건이었다. 사람인과 잡코리아의 경우 민간 구인구직사이트 경쟁사로서 상호 동시 게시된 구인정보는 거의 없었다. 따라서 동시에 3개 웹사이트에 동시에 게시된 구인정보 분석이 불가능하였다.

워크넷, 잡코리아 동시 게시 구인정보를 필터링하기 위해 데이터베이스 연결 형태와 사용한 쿼리는 다음 <Figure 4>와 같다.

필터링 결과 2개 웹사이트에 동시에 게시된 구인정보의 레코드 수는 11건이었다. 워크넷에 구인정보를 게시할 경우 잡코리아에서는 별도로 게시하지 않고 워크넷에서 가져와 게시하는 경우가 대부분이며, 반대방향으로도 동일했다. 따라서 워크넷과 잡코리아에 동시에 게시된 구인정보가 분석할 만큼 충분하지 않았다.



<Figure 3> Explore job postings on three websites simultaneously



<Figure 4> Explore job postings on WORKNET and JOBKOREA simultaneously

워크넷, 잡코리아 동시 게시 구인정보를 필터링하기 위해 데이터베이스 연결 형태와 사용한 쿼리는 다음 <Figure 5>와 같다.

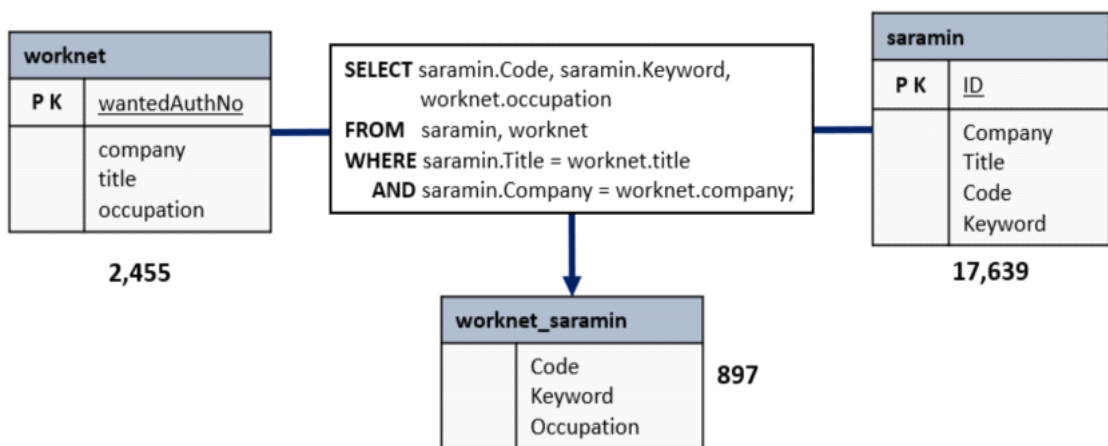
필터링 결과 2개 웹사이트에 동시에 게시된 구인정보의 레코드 수는 897건이었다. 사람의 경우 타 사이트에 게시된 구인정보를 자기 사이트로 가져와 게시하지 않고 별도로 자기 사이트에 입력해 게시하는 경우가 많았다. 따라서 워크넷과 사람인에 동시에 게시된 구인정보가 분석할 만큼 수집 가능했다.

웹사이트 상에서 통합적으로 검색이 되는 것과 API로 채용정보 데이터를 제공하는 것은 상당히 달랐다. 연결되는 채용정보가 많을 것으로 예상했으나 특정 데이터는 API를 통해 제공되지 않게 되면서 자동 수집 및 매칭 중 일부만 가능한 상황이 발생했다.

워크넷, 잡코리아, 사람인에서 데이터를 공유하는 방식을 살펴보면, 워크넷은 다른 곳으로부터 온 데이터는 오픈 API를 통해 제공하지도 않을 뿐만 아니라 자체 분류체계로 환원시키지 않고, 본래의 분류체계와 양식을 그대로 따라서 제

공했다. 잡코리아는 워크넷으로부터 채용공고를 가져갈 때 그것을 자체 분류에 맞게 변환하는 규칙을 보유하고 있었다. 외부 정보제공처의 데이터에 대해, 시스템에서 내부적인 맵핑을 수행하고 있었다. 외부의 데이터는 잡코리아의 분류코드를 부여받아 사전 연구에서 발견했던 것처럼 분석의 대상이 될 수 있었다. 그러나 API를 통해 외부 데이터를 제공하지 않기 때문에 수작업으로 수집하지 않는 이상 데이터 워크넷-잡코리아 간 연계가 불가능했다. 사람은 외부의 데이터라도 요청을 통해 자체 데이터로 편입시켜 분류하고, 그 데이터를 API로 제공하기 때문에 분석 대상으로 적합했다.

따라서 최종적으로 워크넷-사람인 간 매칭 및 연관분석을 수행했다. 워크넷-잡코리아 간 매칭은 잡코리아 측에 요청하여 내부적인 맵핑 규칙을 획득하고 그것을 통해 통찰을 얻기로 결정했다. 워크넷-사람인 간의 데이터는 약 900건이므로, 충분히 분석할 만한 가치가 있다고 판단하여 연관분석을 수행하고자 한다.



<Figure 5> Explore job postings on WORKNET and saramin simultaneously

4.3 직종 간 연관분석

워크넷과 사람인에서 동시 발생한 총 897개의 구인정보를 이용하여 직종 간 연관분석을 실시하였다. ‘빈발 패턴 마이닝(frequent pattern mining)’인 ‘Apriori 알고리즘’을 적용하기 위해 R의 연관분석 패키지(arules)를 이용해 발생빈도 1회 이상, 신뢰도 90% 이상으로 연관규칙을 도출하였다.

빈발 패턴 마이닝은 장바구니 분석(market basket analysis), 연관규칙분석(association rule mining)이라고도 불리는 데이터마이닝 기법으로 원래 동시에 구매하는 물건들 간의 규칙을 찾기 위해 고안되었다(Kim et al., 2012). 아이템은 장바구니 분석에서는 물건이며, 트랜잭션을 구성하는 하나의 개체이고, 트랜잭션은 한 건의 거래를 의미하며, 여러 개의 아이템을 포함하여 해당하는 아이템들이 ‘장바구니에 동시에 담겼다’, 즉, 동시에 발생하였다는 의미를 내포한다. 빈발 패턴 마이닝의 입력(input)은 일련의 트랜잭션이며, 출력(output)은 트랜잭션들 속에서 발견되는 규칙이다.

규칙의 평가지표로 지지도(support)와 신뢰도(confidence)가 있다. 지지도(support)는 전체에서 규칙을 만족하는 경우가 몇 건 등장했는지를 나타내는 빈도이고, 신뢰도는 규칙의 조건절 부분이 발생한 건수 중에서 결과절 부분이 발생한 건수로서 상대적인 개념의 빈도이다. 본 연구의 데이터 특성상 지지도보다는 신뢰도가 더 좋은 규칙 판단 기준일 것으로 보인다. 전체적인 데이터의 분포, 즉, 분류별 채용공고의 수가 비슷한 수준이라면 지지도가 의미를 가질 수 있으나, 분류별 채용공고의 수가 큰 차이가 있기 때문에 해당하는 분류 내에서 규칙을 평가할 수 있도록 신뢰도를 주요 기준으로 채택했다.

Apriori 알고리즘은 ‘더 긴 연관규칙은 더 짧은 연관규칙으로부터 나온다’라는 Apriori 규칙에

근거하여 연관규칙을 찾아내는 빈발 패턴 마이닝 알고리즘이다.

입력 데이터를 구성하기 위해 빈발 패턴 마이닝에서는 일련의 트랜잭션을 입력받지만, 여기에서는 같은 채용공고에 대한 여러 분류를 하나의 트랜잭션으로 취급했다. 즉 하나의 아이템은 각 채용공고 사이트의 직종코드이며, 하나의 트랜잭션 내에서 동시 발생했다고 볼 수 있다. 하나의 동일한 채용공고가 동시에 여러 분류로 나눠졌다고 볼 수 있다.

추출된 규칙 중 일부는 <Table 2>와 같다. 규칙은 코드 체계로 되어 있으나, 매칭을 통해 해당하는 분류의 이름과 연결시킬 수 있다. 088504는 멀티미디어 디자이너이며, 첫 번째 규칙은 {멀티미디어디자이너(085504)} => {웹디자인}으로 바꿀 수 있다. 즉 워크넷 분류 [멀티미디어디자이너]라면, 사람의 키워드 [웹디자인]을 가진다는 의미한다.

<Table 2> Association rules

No	Rule	Support
1	{085504} → {Web design}	0.012263
2	{085504} → {085500}	0.012263
3	{085504} → {412 }	0.012263
4	{System Analysis} → {409}	0.010033
5	{System Analysis} → {202100}	0.010033
6	{Bulletin Board Management} → {406}	0.011148
...

동시에 게시된 구인정보가 속한 직종을 연결함으로써 각 구인구직사이트의 직종 간 연관분석 실시했다. 워크넷의 최하위 분류 직종인 3차 직종 55개와 사람인의 최하위 분류 직종인 2차 직종 16개 간의 연관규칙을 도출했다. 양방향 2개의 규칙 테이블의 예는 다음 <Table 3>과 같다.

(Table 3) Example of association rule table between WORKNET 3rd level and saramin 2nd level

WORKNET → saramin	saramin → WORKNET
System SW developer (203100) → System programmer (408)	DB · DBA (416) → Application SW developer (203200)
System SW and OS programmer(203101) → HW · SW (411)	Web planning · web marketing · PM(403) → Web and multimedia planner (204200)
Webmaster and webPD (204101) → Webmaster · QA · tester (401)	Web master · QA · tester (401) → Webmaster and web developer (204100)
...	

또한 워크넷의 최하위 분류 직종인 3차 직종 55개와 사람인의 최하위 분류 직종 아래의 키워드 243개와 연관규칙 도출하였다. 양방향 2개의 규칙 테이블의 예는 다음 <Table 4>과 같다.

(Table 4) Example of association rule table between WORKNET 3rd level and saramin the lowest keyword level

WORKNET → saramin	saramin → WORKNET
Communication technology developer and network operator(201203) → Communication	Bulletin board management → Web and multimedia designers (085500)
Webmaster and webPD (204101) → Webmaster	Embedded → System SW developer (203100)
IT consultant (202101) → Consultant	Electronic commerce → Web and multimedia designers (085500)
...	...

4.4 직무분류체계 맵핑

직종 간 연관분석으로 도출한 연관규칙을 이용해 워크넷과 사람인의 직무분류체계를 맵핑하였다. <Figure 6>은 전체적인 연결을 반영한 최종 맵핑의 형태이며, 최종 맵핑은 각각의 세부 분류 간 맵핑 관계에 기반을 두고 분석되었다.

사람인 401 분류(웹마스터 · QA · 테스터)에 대한 워크넷 맵핑을 살펴보면, 워크넷 웹마스터 및 웹PD(204101)은 사람인 웹마스터 · QA · 테스터(401)으로 연결되었다. 또한 사람인 402 분류(서버 · 네트워크 · 보안)에 대한 워크넷 맵핑을 살펴보면, 워크넷 네트워크시스템 분석가 및 개발자(운영자 제외)(202200), 컴퓨터시스템 및 네트워크 보안전문가(202300), 네트워크프로그래머(203205), 정보시스템운영자(205200), 서버운영관리자(205201), 네트워크운영관리자(클라우드컴퓨터운영관리자 포함)(205202), IT기술지원전문가(205203), 컴퓨터시스템 설계 및 분석가(202102)는 사람인 서버 · 네트워크 · 보안(402)으로 연결되었다.

다음으로 구인구직사이트의 직무분류체계 맵핑에서 허브 역할을 했던 워크넷의 직무분류체계와 SQF의 직무분류체계를 맵핑했다. SQF 기반 구인정보를 필터링하여 워크넷의 최하위 분류인 3차직종 55개와 SQF의 최하위 분류인 세분류 22개를 수작업으로 맵핑했다. 단 SQF 대신 NCS 기반 구인정보를 조사하였으나 10개 이하로 극히 드물어 데이터가 없는 최하위 분류의 경우 워크넷 사람인 간의 연관규칙을 참고해 수작업으로 맵핑했다.

구인구직사이트의 직무분류체계 맵핑 결과와 SQF 직무분류체계 맵핑 결과를 토대로 시장수요 중심의 새로운 표준 직무분류체계를 <Figure 7>

2차 분류 코드	3차 분류 코드	워크넷 3차 분류 직종	사람인		
			1:1매칭	다:1매칭	1:다매칭
131. 정보통신 관리자, 컴퓨터 및 통신 공학기술자(엔지니어)	016100	정보통신 관련 관리자(부서장)	N/A		
	201100	컴퓨터하드웨어 기술자 및 연구원		하드웨어 소프트웨어(411)	
	201101	컴퓨터하드웨어 기술자 및 연구원(컴퓨터 부품 및 주변기기 제외)		하드웨어 소프트웨어(411)	
	201102	컴퓨터 부품 및 주변기기 개발기술자 및 연구원		하드웨어 소프트웨어(411)	
	201200	통신공학 기술자 및 연구원	통신 모바일(410)		
	201201	통신모바일기술자 및 연구원		통신 모바일(410)	
	201202	통신기기 및 장비 개발자 및 연구원(무선 등 제외)		통신 모바일(410)	
	201203	통신기술개발자 및 통신망운영기술자		통신 모바일(410)	
	201204	통신관리기술자		통신 모바일(410)	
	202100	정보시스템 설계 및 분석가			ERP-시스템분석 설계(409)
	202101	IT컨설턴트	컨설턴트		ERP-시스템분석 설계(409)
202102	컴퓨터시스템 설계 및 분석가				
202103	컴퓨터시스템관리기술자	N/A			
202200	네트워크시스템 분석가 및 개발자(운영자 제외)		서버 네트워크 보안(402)		
202300	컴퓨터시스템 및 네트워크 보안전문가		서버 네트워크 보안(402)		
203100	시스템소프트웨어개발자	시스템프로그래머(406)			
133. 소프트웨어 개발, 프로그래머	203101	시스템소프트웨어 및 운영체제 프로그래머	시스템 프로그래머	시스템프로그래머(408)	
	203102	임베디드 및 임베디드 프로그래머		시스템프로그래머(408)	임베디드
	203102	임베디드 및 임베디드 프로그래머		시스템프로그래머(408)	필웨어
	203200	응용소프트웨어개발자(응용프로그래머)	응용프로그래머(407)		
	203201	C프로그래밍언어 전문가	C++	응용프로그래머(407)	
	203202	JAVA프로그래밍언어 전문가	Java	응용프로그래머(407)	
	203203	그 외 프로그래밍언어 전문가		응용프로그래머(407)	
	203204	응용소프트웨어프로그래머(네트워크 게임 제외)	삭제		
	203205	네트워크프로그래머	네트워크엔지니어	서버 네트워크 보안(402)	
	203206	컴퓨터 및 모바일 게임프로그래머	게임 Game(405)		
	134. 웹기획, 웹프로그래머, 웹디자인, 게임	085500	웹 및 멀티미디어 디자이너	웹디자인(412)	
085501		웹디자이너	웹디자인(412)		
085502		멀티미디어디자이너(게임분야)		웹디자인(412)	
085503		멀티미디어디자이너(방송영상분야)		웹디자인(412)	
085504		멀티미디어디자이너(게임 방송영상 외)		웹디자인(412)	
203206		컴퓨터 및 모바일 게임프로그래머	게임 Game(405)		
204100		웹마스터 및 웹개발자(웹프로그래머)			웹마스터 OA 테스터(401)
204100		웹마스터 및 웹개발자(웹프로그래머)			웹프로그래머(404)
204101		웹마스터 및 웹PD	웹마스터 OA 테스터(401)		
204102		웹개발자(웹프로그래머 및 웹엔지니어)	웹프로그래머(404)		
204200		웹 및 멀티미디어 기획자		웹기획 웹마케팅 PM(403)	
204201	웹기획자		웹기획 웹마케팅 PM(403)		
204202	멀티미디어기획자				
204203	컴퓨터 및 모바일 게임기획자	게임기획(405)			
205100	데이터베이스 설계 구축 및 관리 기술자	데이터베이스 DBA(416)			
135. 데이터베이스, 시스템 운영	205200	정보시스템운영자	시스템운영		
	205201	서버운영관리자	서버관리	서버 네트워크 보안(402)	
	205202	네트워크운영관리자(클라우드컴퓨팅운영관리자 포함)	네트워크관리	서버 네트워크 보안(402)	
	205203	IT기술지원전문가		서버 네트워크 보안(402)	

<Figure 6> Mapping between WORKNET and saramin Job Classification System

과 같이 제안했다. 기존의 다른 분류체계와의 차이점은 분류의 단계(depth)가 상대적으로 유연하다는 점에 있다. 워크넷은 직무로만 3차 분류까지 나누어 두었고, 잡코리아는 직무로만 2차 분류까지 나누어 두었으며, 그 이하는 키워드 형태로 직무를 세분화했다. 사람인은 직무로만 2차 분류까지 나누어 두었으며, 그 이하는 키워드 형태로 직무를 세분화하였다. 본 연구에서 새롭게 제안한 표준 직무분류체계는 키워드 기반의 직무를 일부 수용하였으며, 제품명이 들어간 경우도 직무로 취급하였다. 본 분류체계에서는 2차 분류에서 끊어진 직무도 있을 뿐 아니라 4차 분

류까지 세분화된 직무도 있어, 모든 직무가 동일한 단계로 세분화될 수 없다는 아이디어를 반영했다. 또한 시장 데이터를 연관분석하여 나온 연관규칙과 전문가의 견해를 합쳐 제안한 것이기 때문에 기존 방법론을 통한 직무분류체계와는 달리 시장 수요를 반영한 데이터 기반의 지능형 직무분류체계라고 볼 수 있다.

<Figure 7>에서 각 색깔에 따라 [검정: 워크넷], [녹색: SQF], [보라색: 사람인], [하늘색: 잡코리아], [빨강색: 자체수정]을 의미한다. 새로운 직무 분류 체계에서 1차 분류 IT컨설팅, 컴퓨터 시스템, 네트워크, 보안과 관련된 직무 체계는 2

순번	1차 분류	2차 분류	3차 분류	4차 분류	빈도수	비고
1	정보통신 관리, 컴퓨터 및 통신 공학기술	IT프로젝트관리			9	
2	정보통신 관리, 컴퓨터 및 통신 공학기술	컴퓨터하드웨어 기술 및 연구	컴퓨터하드웨어 기술 및 연구(컴퓨터 부품 및 주변기기 제외)			
3	정보통신 관리, 컴퓨터 및 통신 공학기술	컴퓨터하드웨어 기술 및 연구	컴퓨터 부품 및 주변기기 개발기술 및 연구			
4	정보통신 관리, 컴퓨터 및 통신 공학기술	통신공학 기술 및 연구	휴대폰스마트폰 개발 및 연구			
5	정보통신 관리, 컴퓨터 및 통신 공학기술	통신공학 기술 및 연구	통신기기 및 장비 개발 및 연구(휴대폰 등 제외)			
6	정보통신 관리, 컴퓨터 및 통신 공학기술	통신공학 기술 및 연구	통신기술개발 및 통신망운영기술			
7	정보통신 관리, 컴퓨터 및 통신 공학기술	통신공학 기술 및 연구	통신감리기술			
8	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	정보시스템 설계 및 분석	엔터프라이즈 시스템 컨설팅			
9	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	정보시스템 설계 및 분석	정보기술전략-계획	정보기술전략	815	정보기술전략-계획 빈도수 합
10	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	정보시스템 설계 및 분석	정보기술전략-계획	정보기술컨설팅	815	정보기술전략-계획 빈도수 합
11	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	정보시스템 설계 및 분석	정보기술전략-계획	정보기술기획	815	정보기술전략-계획 빈도수 합
12	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	정보시스템 설계 및 분석	정보기술전략-계획	SW재용기획	815	정보기술전략-계획 빈도수 합
13	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	정보시스템 설계 및 분석	정보기술전략-계획	빅데이터분석		
14	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	정보시스템 설계 및 분석	정보기술전략-계획	빅데이터분석		
15	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	정보시스템 설계 및 분석	정보시스템 설계 및 분석	정보시스템	3	
16	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	정보시스템 설계 및 분석	정보시스템 설계 및 분석	정보시스템		
17	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	NW엔지니어링			5132	
18	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	NW엔지니어링			5132	
19	IT컨설팅, 컴퓨터시스템, 네트워크, 보안	보안엔지니어링			1230	
19	소프트웨어 개발, 프로그램	SW아키텍처			345	
20	소프트웨어 개발, 프로그램	시스템소프트웨어개발	시스템SW엔지니어링		1529	
21	소프트웨어 개발, 프로그램	시스템소프트웨어개발	임베디드SW엔지니어링	임베디드 시스템 프로그래밍	2278	임베디드SW엔지니어링 빈도수 합
22	소프트웨어 개발, 프로그램	시스템소프트웨어개발	임베디드SW엔지니어링	동역어 시스템 프로그래밍	2278	임베디드SW엔지니어링 빈도수 합
23	소프트웨어 개발, 프로그램	응용SW엔지니어링	C프로그래밍		2863	응용SW엔지니어링 빈도수 합
24	소프트웨어 개발, 프로그램	응용SW엔지니어링	JAVA프로그래밍		2863	응용SW엔지니어링 빈도수 합
25	소프트웨어 개발, 프로그램	응용SW엔지니어링	그 외 프로그래밍		2863	응용SW엔지니어링 빈도수 합
26	웹기획, 웹프로그래밍, 멀티자인, 게임	UI/UX엔지니어링	웹디자인		14560	
27	웹기획, 웹프로그래밍, 멀티자인, 게임	UI/UX엔지니어링	멀티미디어디자인(게임분야)		2649	UI/UX엔지니어링 빈도수 합
28	웹기획, 웹프로그래밍, 멀티자인, 게임	UI/UX엔지니어링	멀티미디어디자인(방송영상분야)		2649	UI/UX엔지니어링 빈도수 합
29	웹기획, 웹프로그래밍, 멀티자인, 게임	UI/UX엔지니어링	멀티미디어디자인(게임, 방송영상 외)		2649	UI/UX엔지니어링 빈도수 합
30	웹기획, 웹프로그래밍, 멀티자인, 게임	컴퓨터 및 모바일 게임프로그래밍				
31	웹기획, 웹프로그래밍, 멀티자인, 게임	웹마스터 및 웹개발(웹프로그래밍)	웹마스터 및 웹PD			
32	웹기획, 웹프로그래밍, 멀티자인, 게임	웹마스터 및 웹개발(웹프로그래밍)	웹개발(웹프로그래밍 및 웹엔지니어링)	웹프로그래밍		
33	웹기획, 웹프로그래밍, 멀티자인, 게임	웹마스터 및 웹개발(웹프로그래밍)	웹개발(웹프로그래밍 및 웹엔지니어링)	HTML,요청-컨텐츠관리		
34	웹기획, 웹프로그래밍, 멀티자인, 게임	웹 및 멀티미디어 기획	웹기획			
35	웹기획, 웹프로그래밍, 멀티자인, 게임	웹 및 멀티미디어 기획	멀티미디어 기획			
36	웹기획, 웹프로그래밍, 멀티자인, 게임	웹 및 멀티미디어 기획	컴퓨터 및 모바일 게임기획			
37	데이터베이스, 시스템 운영	DB엔지니어링			1259	
38	데이터베이스, 시스템 운영	IT시스템관리	웹사이트운영		1230	IT시스템관리 빈도수 합
39	데이터베이스, 시스템 운영	IT시스템관리	서버운영관리		1230	IT시스템관리 빈도수 합
40	데이터베이스, 시스템 운영	IT시스템관리	네트워크운영관리(클라우드컴퓨터운영관리 포함)		1230	IT시스템관리 빈도수 합
41	데이터베이스, 시스템 운영	IT기술교육 및 지원			2218	
41	데이터베이스, 시스템 운영	IT기술교육 및 지원			2218	

(Figure 7) New job classification system proposal

차 분류 3개, 3차 분류 5개, 4차 분류 5개로 구성되고, 1차 분류 데이터베이스, 시스템 운영과 관련된 직무 체계는 2차 분류 3개, 3차 분류 3개, 4차 분류 0개로 구성되었다. 1차 분류 웹기획, 웹프로그래밍, 웹디자인, 게임과 관련된 직무 체계는 2차 분류 4개, 3차 분류 9개, 4차 분류 2개로 구성되고, 1차 분류 정보통신 관리, 컴퓨터 및 통신 공학기술과 관련된 직무 체계는 2차 분류 3개, 3차 분류 6개로 구성되었다.

5. 결론 및 시사점

본 연구는 주요 구인구직사이트의 직무분류체계를 맵핑하고 SQF와 직무분류체계를 매핑하여 시장 중심의 새로운 직무분류체계를 제안했다. 구인구직사이트 및 SQF 간의 모든 연계가 수행되었고 그 관계를 기반으로 전문가들이 참여하여 새로운 직무분류체계를 제안하였다.

직종 간 맵핑은 과거 일부 전문가의 수차례 시

도에도 불구하고 최종적으로 합의에 실패했을 뿐만 아니라 NCS는 의무적으로 사용해야 하는 공공기관을 제외한 민간 기업에서는 거의 사용하지 않았다. 본 연구는 일부 전문가의 직관이 아닌 연관분석을 통한 데이터 기반 지능형 직종 맵핑 테이블을 활용하여 시장 중심의 새로운 직무분류체계 제안함으로써 향후 SW산업 분야에 대한 SQF의 수정방향을 제시하였다.

다만 본 연구는 수집 시점이 일시적이기 때문에 시간에 따라 변화하는 시장 수요를 충분히 반영하지 못하는 한계가 있다. 시간의 흐름에 따라 시장에서 요구하는 직종분류체계는 변해갈 것이기에 종단연구가 필요하다. 계절 및 주요 공채 시기 등에 따른 고용 시장의 변화가 발생할 가능성이 높기에 더욱 정확한 매칭을 얻기 위해서는 지속적인 데이터 모니터링과 여러 번에 걸친 실험이 필요하다. 또한 분석 시점을 기준으로 주요 구인구직사이트에 동시 게시된 구인정보가 충분하지 않아, 일부 직무 간 맵핑의 경우 연관분석을 할 수 없었다. 일부 수행 가능했던 연관분석의 결과에 기반을 두고 분석 전문가들이 참여하여 보완하였으나, 추가적인 데이터 수집을 통한 연관분석을 수행할 필요성이 있다. 더불어 직무 간 연관분석 결과와 직무분류체계 맵핑을 통해 제안한 시장 중심의 직무분류체계 결과에 대해 업계 실무자 또는 전문가의 검증이 필요하다.

본 연구의 결과를 활용하여 워크넷, 잡코리아, 사람인 등과 같이 채용 관련 정보를 제공하는 사이트와 연동한다면 구직, 구인, 채용중계 정보의 활용을 증대시킬 수 있다. 구인기업, 구직자뿐만 아니라 채용중계업자까지 모두 활용 가능한 표준 직무분류체계로서 기능할 것이다. 채용지원 서비스를 SW분야의 대표적인 채용지원 체계로 자리매김하는 데 기여하고, 자가진단 및 역량 상

승의 기준으로서 활용 가능하다. 본 연구 결과를 바탕으로 향후 SW산업 분야에 대한 SQF의 수정 방향을 제시할 수 있고, SW산업 분야에서의 경험은 타 산업 분야로 이전되어 확장할 수 있을 것으로 기대한다.

본 연구 결과를 통해 SW기업은 우수 고급인력을 원활하게 확보하고 능력 중심의 채용 문화를 조성하여 잡 미스매칭을 해소하고 SW 고급인재를 쉽게 확보할 수 있을 것이다. 직무능력 중심의 NCS를 기반으로 한 SW분야 SQF와 실제 산업 현장과의 괴리를 줄여 체계적이고 현실적으로 인력을 공급할 수 있는 서비스 활성화 방안을 마련하였다. 실제 산업 현장에서 적용 가능한 SQF를 통해 신입자와 경력자는 희망하는 일자리를 찾고 기업은 원하는 인재를 확보할 수 있다. SW 전문 인력을 양성하고 SW분야에 인재를 유입시켜 국내 SW산업을 성장시키고 이를 다양한 산업에 전파하여 국가경쟁력을 강화시킬 수 있다.

본 연구는 일부 전문가의 직관이 아닌 직종 간 연관 분석을 통해 데이터 기반 직종 맵핑을 제안하고 시장 중심의 새로운 직무분류체계를 제안하여 직무분류체계 분야에서도 일부 전문가의 직관이 아닌 과학적인 데이터 분석을 통해 향후 표준 직무분류체계 제안이 가능해 질 것으로 기대한다.

참고문헌(References)

- Ahn, I., "A Study of Inter - occupational Relationship in Job Analysis and Vocational Trend in Information Management and Service," *Journal of the Korean Biblia*

- Society for Library and Information Science*, Vol.16, No.2(2005), 225~245
- IPA(Information-Technology Promotion Agency), “T스킬標準 V 3 (バージョン 3) について,” *Technical Report*, 2009
- Jun, H.-J., T.-S. Kim, J. Yoo, and S.-H. Gee, “Development of Skills Framework for Information Security Workforce,” *Journal of the Korea Institute of Information Security & Cryptology*, Vol.19, No.3(2009), 143~152.
- Kang, E.-S. and D.-S. Ko, “Automatic Classification Model of Electronic Documents Based on Machine Learning for Job Analysis,” *The Journal of Korean Institute of Information Technology*, Vol.17, No.7(2019), 23~29.
- Kim, M., N. Kim, and J.-H. Ahn, “An Investigation on Expanding Co-occurrence Criteria in Association Rule Mining,” *Journal of Intelligence and Information Systems*, Vol.18, No.1(2012), 23~38
- Na, S. H., “Policy Suggestions for SW Manpower Training,” *Technical Report*, K, Korea Software Industry Association, 2013
- NWCET(National Workforce Center for Emerging Technologies), “Building a Foundation for Tomorrow: Skill Standards for Information Technology,” *Technical Report*, 2003
- Park, H. S. et al., “How to build a Sectoral Qualifications Framework(SQF) in SW industry to build a National Qualifications Framework(NQF),” *Technical Report*, K, Korea Information Society Development Institute, 2016
- SFIA(Skills Framework for the Information Age), “Skills framework for the information age, version. 3,” *Technical Report*, 2005

Abstract

Development of Intelligent Job Classification System based on Job Posting on Job Sites

Jung Seung Lee*

The job classification system of major job sites differs from site to site and is different from the job classification system of the ‘SQF(Sectoral Qualifications Framework)’ proposed by the SW field. Therefore, a new job classification system is needed for SW companies, SW job seekers, and job sites to understand. The purpose of this study is to establish a standard job classification system that reflects market demand by analyzing SQF based on job offer information of major job sites and the NCS(National Competency Standards).

For this purpose, the association analysis between occupations of major job sites is conducted and the association rule between SQF and occupation is conducted to derive the association rule between occupations. Using this association rule, we proposed an intelligent job classification system based on data mapping the job classification system of major job sites and SQF and job classification system. First, major job sites are selected to obtain information on the job classification system of the SW market. Then We identify ways to collect job information from each site and collect data through open API. Focusing on the relationship between the data, filtering only the job information posted on each job site at the same time, other job information is deleted. Next, we will map the job classification system between job sites using the association rules derived from the association analysis. We will complete the mapping between these market segments, discuss with the experts, further map the SQF, and finally propose a new job classification system.

As a result, more than 30,000 job listings were collected in XML format using open API in 'WORKNET,' 'JOBKOREA,' and 'saramin', which are the main job sites in Korea. After filtering out about 900 job postings simultaneously posted on multiple job sites, 800 association rules were derived by applying the Apriori algorithm, which is a frequent pattern mining. Based on 800 related rules, the job classification system of WORKNET, JOBKOREA, and saramin and the SQF job classification system were

* Corresponding Author: Jung Seung Lee
School of Business, Hoseo University
12, Hoseodae-gil, Dongnam-gu, Cheonan-si, Chungcheongnam-do, 31066, Korea
Tel: +82-41-560-8361, Fax: +82-41-560-8307, E-mail: leebahk@gmail.com

mapped and classified into 1st and 4th stages.

In the new job taxonomy, the first primary class, IT consulting, computer system, network, and security related job system, consisted of three secondary classifications, five tertiary classifications, and five fourth classifications. The second primary classification, the database and the job system related to system operation, consisted of three secondary classifications, three tertiary classifications, and four fourth classifications. The third primary category, Web Planning, Web Programming, Web Design, and Game, was composed of four secondary classifications, nine tertiary classifications, and two fourth classifications. The last primary classification, job systems related to ICT management, computer and communication engineering technology, consisted of three secondary classifications and six tertiary classifications. In particular, the new job classification system has a relatively flexible stage of classification, unlike other existing classification systems. WORKNET divides jobs into third categories, JOBKOREA divides jobs into second categories, and the subdivided jobs into keywords. saramin divided the job into the second classification, and the subdivided the job into keyword form. The newly proposed standard job classification system accepts some keyword-based jobs, and treats some product names as jobs. In the classification system, not only are jobs suspended in the second classification, but there are also jobs that are subdivided into the fourth classification. This reflected the idea that not all jobs could be broken down into the same steps. We also proposed a combination of rules and experts' opinions from market data collected and conducted associative analysis. Therefore, the newly proposed job classification system can be regarded as a data-based intelligent job classification system that reflects the market demand, unlike the existing job classification system.

This study is meaningful in that it suggests a new job classification system that reflects market demand by attempting mapping between occupations based on data through the association analysis between occupations rather than intuition of some experts. However, this study has a limitation in that it cannot fully reflect the market demand that changes over time because the data collection point is temporary. As market demands change over time, including seasonal factors and major corporate public recruitment timings, continuous data monitoring and repeated experiments are needed to achieve more accurate matching. The results of this study can be used to suggest the direction of improvement of SQF in the SW industry in the future, and it is expected to be transferred to other industries with the experience of success in the SW industry.

Key Words : Association rule, Frequent pattern mining, Job classification system, Software Industry

Received : November 13, 2019 Revised : December 22, 2019 Accepted : December 23, 2019

Publication Type : Regular Paper Corresponding Author : Jung Seung Lee

저 자 소개



이정승

현재 호서대학교 경영학부 부교수 및 스마트경영 연구센터 센터장으로 재직 중이다. KAIST 경영과학과에서 학사와 석사학위를 취득하고, 동 대학 테크노경영대학원에서 지능정보시스템을 전공으로 경영공학전공 박사학위를 취득하였다. 한국지능정보시스템학회 총무부회장, JITAM(Journal of Information Technology Applications and Management) 편집위원장, 한국데이터전략학회 부회장을 역임하였다. 주요 관심분야는 인공지능 응용, 스마트 그리드, 지속가능 공급망 등이다.