

논문 2019-56-10-6

공공 데이터를 활용한 대학생 진로 조언을 위한 그래프데이터베이스 기반 지식베이스 구축

(Construction of Knowledge Base Based on Graph Database for
College Student Career Advice Using Public Data)

김 중 율*, 노 광 현**

(Jong-Yul Kim and Kwang-Hyun Ro[©])

요 약

본 연구는 대학생들의 진로 문제 해결을 위한 그래프데이터베이스를 활용한 진로조언 시스템 구축 연구로 직업 및 진로관련 공공 데이터에서 수집된 진로 정보들의 연관 관계를 설정하여 그래프데이터베이스 기반의 지식베이스를 구축하고, 구축된 지식베이스를 활용하여 개별 학생의 e-포트폴리오 데이터와의 매핑으로 직업 추천, 역량 개발 정보 등의 다양한 진로 정보를 학생들에게 제공해 주기 위한 진로조언 시스템 모델 연구이다. 진로 조언을 위한 데이터로 고용정보원의 직무데이터를 수집하여 직무 능력 단위에 대한 지식 및 기술, 수행태도, 관련자격증, 관련전공, 관련훈련 요소 간의 연관 관계를 분석·설계하여 그래프데이터베이스 기반의 지식베이스를 구축하고 가상의 e-포트폴리오 정보와 매핑하여 대학생 맞춤형 진로조언 모델을 실제 구현하고 시나리오에 따른 직무 추천과 대화형 인터페이스를 활용한 진로 조언 모델 검증과 활용 가능성을 제시하였다.

Abstract

This study is to build a career advice system using a graph database to solve career problems of college students. It builds a knowledge base based on the graph database by establishing a relationship between career information collected from occupational and career-related public data. This study is a career advice system model to advise students of various career information such as job recommendation and competency development information by mapping with the ePortfolio data of individual students using knowledge base. Collects job data API of employment information source as data for career advice, analyzes and designs the knowledge and skills of job competency unit, performance attitude, related qualifications, related majors, and related training elements. By constructing and mapping with virtual e-portfolio information, this study presented the possibility of realizing college career customized career advice model and using career advice through scenario recommendation and interactive interface.

Keywords: Career advisor, Graph database, Knowledge Base, Bigdata, e-Advisor

I. 서 론

2017년 교육부의 ‘대학생진로교육현황조사’에 따르면 국내 대학생들의 가장 큰 고민은 진로 및 학업으로 조

사되었으며^[1], 동년 중앙대에서 실시한 ‘취업·진로 의식 설문조사’에서는 자신의 능력과 역량, 흥미와 적성에 대해 잘 모르겠다는 응답이 54%, 직업의 종류와 세계에 대해 잘 모르겠다는 응답이 18%, 목표를 달성하기 위한 노력과 의지가 부족하다는 응답이 14%로 조사된 것과 같이 대다수의 학생이 진로 정보의 부족과 목표 설정이 부족한 것으로 확인되었다. 목표가 정해지면 그 목표 실천을 위해 적극적으로 노력하게 되며^[2], 자기주도적인 심리 변인이 진로결정 자기효능감과 진로준비행동에 긍정적 영향^[3]을 줄 수 있기에 ‘대학생진로교육현황조사’ 응답 결과와 같이 대학생들이 필요로 하는 전공 및 실

* 학생회원, ** 정회원, 한성대학교 IT융합공학부
(Division of IT Convergence Engineering, Hansung University)

© Corresponding Author(E-mail : khrho@hansung.ac.kr)

※ This research was financially supported by Hansung University

Received ; June 3, 2019 Revised ; September 16, 2019

Accepted ; September 23, 2019

무 중심의 전문화된 진로 지도와 진로·취업 정보가 대학생들에게 가장 지원이 필요한 분야라 할 수 있다. 이러한 진로 문제는 대학생 개개인의 문제뿐만이 아니라 청년층의 실업 증가와 중소기업 등에서의 인력수급 미스매치 문제와 함께 공무원시험준비생의 증가와 같은 다양한 사회 문제 원인 중 일부가 되기도 한다^[4~5]. 이러한 대학생들의 진로문제 해결을 위해 대학들 역시 학생들의 진로지도를 위한 진로지원센터를 신설하거나 진로지도를 위한 다양한 프로그램들을 운영하고 있으나 대면 상담의 부담감과 학생들의 눈높이에 맞지 않는 진로지도 등의 이유로 대부분의 학생에게 외면받고 있는 것으로 조사되었다^[1].

미국, 독일 등의 해외 대학의 경우 대학의 다양한 문제들에 대해 전통적인 정보 기술에서 구현이 어렵거나 불가능했던 학생 개인의 맞춤형 코스설계, 전공선택, 수강신청과 같은 문제들을 빅데이터 분석, 인공지능과 같은 현대 정보 기술을 활용하여 해결하고자 하는 연구가 증가하고 있으며, 높은 성과를 보이고 있다^[6]. 정보기술을 활용한 대학 혁신 사례로 대표적인 빅데이터 기반의 맞춤형 지원시스템인 애리조나주립대학교의 e-Advisor나 Summer-Melt 문제 해결을 위한 조지아주립대학교의 Pounce, 베를린 공대의 챗봇 안내 시스템 Alex와 같이 데이터분석기술과 인공지능기술의 발달은 학생 개인에게 맞춤형 지원과 조언이 가능하게 되었다.

본 연구에서는 국내 대학과 대학생들의 가장 큰 고민인 진로 문제 해결을 위한 지식 데이터로 공공데이터인 한국고용정보원의 직무데이터를 수집하여 직무 데이터 내의 직무 수행을 위한 역량, 지식 및 기술, 수행태도, 관련자격증, 관련전공, 관련훈련 요소와 직무 능력과의 관계를 분석·설계하여 연관 관계를 생성한다. 이러한 연관 관계 분석 및 저장에 최적화된 그래프데이터베이스를 활용하여 지식베이스를 구축하고 사이퍼 쿼리 언어와 대화형 인터페이스를 활용하여 가상의 e-포트폴리오에 등록된 학생의 정보와 매핑하여 맞춤형 직무 추천과 직무를 개발하기 위한 결과 도출을 통해 대학에서 보유하고 있는 학생 데이터 활용 방안과 진로 조언 모델에서의 그래프데이터베이스의 활용 방향 제시를 위한 연구를 수행하였다.

II. 이론적 배경

1. 현 대학에서의 진로지도 문제점

전통적인 직업들과 달리 현대의 직업들은 매우 다양

하게 세분화 되어있고 이렇게 세분화된 직업들은 다시 기술, 지식 등의 수많은 관계와 밀접하게 엮여있다. 그 때문에 대부분의 대학생들은 자신에게 맞는 진로 관련 정보를 찾거나 진로와 연관된 전문가들에게 조언을 구하는 것은 매우 힘든 것이 현실이다. 현 대학들이 도입하고 있는 진로·취업지원센터 역시 정형화된 심리검사나 진로 설계를 위한 프로그램의 수행 이외에 실제로 대학생들이 원하는 전공 맞춤형 진로 지원은 불가능하다. 이러한 문제점으로 인하여 많은 대학생은 진로에 대한 충분한 고민과 정보가 부족한 상태에서 대학을 졸업하고 취업 시장에 내몰리게 되면서 공시생의 증가, 졸업유예, 첫 취업 후 1년 내 퇴사자 증가와 같은 사회 문제의 원인이 되기도 한다^[7]. 공시생들 중 30% 이상 이 진로변경을 희망하고 있는 것과 같이 진로의 문제와 고민은 대학에서만 아니라 대학 졸업 이후 생애 전주기에 걸쳐 계속 이어지고 있다^[8]. 이러한 문제 해결을 위해 정부 역시 다양한 직업훈련과정 개설을 통해 해당 직무 능력 배양 등의 다양한 지원을 하고 있으나 직무 위주의 교육 이외에 학생 개개인의 맞춤형 진로 지원은 매우 미흡한 것이 현실이다.

이러한 문제를 해결하기 위해서는 대학생의 재학 기간에 충분한 직업 및 직무에 대한 정보 제공과 개개인의 특성에 맞는 맞춤형 진로 지도가 필요하다.

2. 대학에서의 정보기술 활용 사례

유에스뉴스앤드월드리포트 선정 2015년부터 2017년까지 가장 개혁적인 대학 1위에 선정된 애리조나 대학교의 e-Advisor는 학생지원부서를 통한 학업 지원의 한계에 대한 문제를 현대정보기술을 활용하여 해결하기 위해 설계되었다. e-Advisor는 대학의 학사 데이터베이스에서 학생의 이력과 경험 등의 데이터를 현대 정보기술을 활용하여 분석하여 복잡한 커리큘럼에서의 체계적인 학생들의 진로를 관리하고 조언할 수 있게 되었다^[9]. 예를 들어 개별 학생의 성과를 예상 성공 패턴과 매핑하여 이수해야 하는 추가 과정을 제안하거나 전공을 변경하는 학생들에게는 이수했던 과정을 분석하여 학위 이수를 위한 최적의 경로를 조언할 수 있다^[6]. e-Advisor는 학생의 입학부터 학위 취득에 이르는 진로(전공) 선택, 학문적 조언, 수강신청, 학생지원 업무 지원과 AI를 활용한 적응형 학습 시스템을 도입으로 기초 수학 통과 학생의 비율이 64%에서 75%로 향상되었으며, 중도 포기 학생의 비율을 16%에서 7%로 낮추는 성과를 보였다^[6,9]. 최근 조지아 공대 Pounce, 베를린 공대 Alex 등의 해외

대학들의 입학부터 졸업까지의 다양한 조언 및 지원 시스템이 연구와 도입이 활발히 진행되고 있으며, 현대 대학생들에게 친숙한 챗봇 형태의 대화형 인터페이스 채택이 증가하고 있다^[10~11].

3. e-포트폴리오

대학에서의 e-포트폴리오는 과거의 학습, 경험, 대외 활동에 대한 학점인정과 포트폴리오를 통하여 학습자 스스로 자신의 학습을 주도해 나갈 수 있도록 지원하는 시스템으로 직업교육 및 훈련 분야에서의 포트폴리오는 학습자 중심, 결과 중심의 평가 목적으로 사용되어 과거의 학습경험에 대한 인증, 자격증, 지속적인 자기계발을 위한 객관적 자료수집에 활용된다^[12]. 대학에 따라 e-포트폴리오는 진로 문제 해결을 위한 접근으로 직업심리검사와 같은 진로설계, 역량 진단과 개발, 수업 이수내역, 수상내역 등의 학생 이력에 대한 내용을 포함하기 때문에 진로 조언 시스템을 설계하기 위한 매칭을 위한 개인화된 데이터로 활용이 가능하다.

4. 진로 조언을 위한 데이터 유형

학생 맞춤형 조언 서비스 제공을 위해서는 직업과 밀접하게 연관된 능력, 자격증, 성격 및 태도, 적성, 전공과 같은 다양한 요소들을 분석하여 지식베이스를 구축하고, 대학생들의 개인화된 데이터를 구축된 지식베이스에 매핑하여 최적화된 정보를 찾을 수 있어야 한다.

이러한 직무와 연관 관계 데이터 수집을 위해서는 공공기관들의 오픈데이터뿐만 아니라 대학, 민간 직업정보제공기업들이 보유하고 있거나 공개하고 있는 다양한 데이터들을 분석하여 직업에 밀접하게 연관되어 있는 다양한 요소들을 분석하고 이를 활용하여 지식베이스를 구축하여야 한다. 다양한 기술의 발전과 더불어 빠르게 변화하고 있는 복잡화된 현대 사회 속에서 신규 직업들의 등장과 직업들과의 연관 요소들 역시 매우 다양하고 서로 간에 밀접하게 연결되고 있다^[13~14].

진로조언시스템을 이용하는 대학생의 e-포트폴리오 내의 개인화된 데이터를 지식베이스에 매핑하여 학생 개개인에게 최적화된 직무를 찾고, 직무개발 방법에 대한 조언을 위해서는 진로 조언에 필요한 최적화된 데이터가 필요하다.

5. 그래프데이터베이스

지식베이스를 구축하는 데 있어 가장 많이 활용되는 RDB(Relational DataBase)의 경우 테이블 조인으로 인

해 2-depth 이상의 관계 데이터를 가져오기 위해서는 복잡한 쿼리와 많은 시간이 소요되기 때문에 다양한 관계를 분석해야하는 지식베이스로 활용하기에는 구조적 한계가 있다. 이러한 소셜네트워크, 교통망, 단백질의 상호작용과 같은 연관 관계 저장과 분석에 대한 최적화된 데이터베이스로 최근 그래프데이터베이스의 연구와 활용이 증가하고 있다.

그래프데이터베이스는 RDB에서 부족했던 관계를 저장하고 탐색하도록 구축된 NoSQL의 종류 중 하나로 관계는 그래프데이터베이스에서 매우 중요한 존재로 데이터베이스의 가치는 대부분 이러한 관계에서 파생된다. 정보를 저장하는 전통적인 데이터베이스와는 달리 그래프에는 정보와 관계, 해당 정보 간의 연결이 있다.

그래프데이터베이스 기반의 지식 그래프는 지식의 영역을 나타내며 실제 개체, 이벤트, 상황과 같은 다양한 개체간의 관계를 체계적으로 연결하게 된다. 즉, 행과 열의 테이블이 아닌 노드와 링크로 구성된 네트워크에 지식을 인코딩하여 데이터의 통합, 지식의 발견 및 심층 분석에 사용하게 되며^[15], IICJ(International Consortium of Investigative Journalists)의 파나마페이퍼 조사에서 조세회피와 같은 숨겨진 관계를 찾아내어 활용성을 인정받았다^[16].

그래프데이터는 데이터 요소인 노드, 두 노드를 연결하는 관계인 엣지(Edge)로 구성되며, 속성에 해당하는 프로퍼티를 가질 수 있다. 예를 들어 그림 1의 ‘공적 개발’이라는 노드는 ‘계량적’, ‘감성적’ 노드와 수행태도라는 엣지 속성을 가지며, 화살표 방향으로 관계가 형성되어 있으며 이러한 노드와 엣지 속성을 빠르고 쉽게 저장하고 조회할 수 있다.



그림 1. 그래프데이터베이스 구조
Fig. 1. Graph Database Structure.

본 논문에서는 Graph DBMS 중 많은 그래프데이터베이스 연구에서 활용되고 있는 GPL3라이선스 오픈소스인 커뮤니티 에디션인 Neo4J를 이용하여 그래프데이터베이스 시스템을 구축하였으며, 그래프데이터베이스에서 효율적인 질의 및 업데이트를 허용하는 선언형 그래프 질의어인 사이퍼(Cypher)를 활용하여 정보의 매핑과 조언시스템에서의 활용 방안을 연구하였다.

III. 데이터 수집 및 진로 조언 시스템 구축

1. 데이터 수집 및 진로 조언 시스템 모델

연구를 위한 시스템 구성 모델은 데이터 수집 및 처리, 그래프데이터베이스, 주기적인 수집을 위한 배치처리 시스템으로 구성하였으며 데이터의 크기와 다양성의 증가에 따라 지식베이스 확장과 연구의 지속성을 고려하여 클라우드 환경으로 시스템을 구축하였다.

가. 연구 시스템 설계

진로조언을 위한 지식베이스 개발을 위한 진로정보 데이터는 그림 2와 같이 직무데이터사전 데이터를 API를 통해 수집하였다. 그러나 이러한 데이터들은 관계가 설정되지 않았기 때문에 불러들인 데이터를 분석하여 데이터 간의 관계를 설정하고 이러한 관계 기반의 그래프데이터베이스에 저장하여 진로조언을 위한 그래프 기반 지식베이스를 구축하였다. 이렇게 구축된 지식베이스는 대학 내부 e-포트폴리오 데이터와 매핑되어 학생들에게 최적의 직무 추천과 해당 직무에 대한 정보를 제공할 수 있도록 설계되었다.

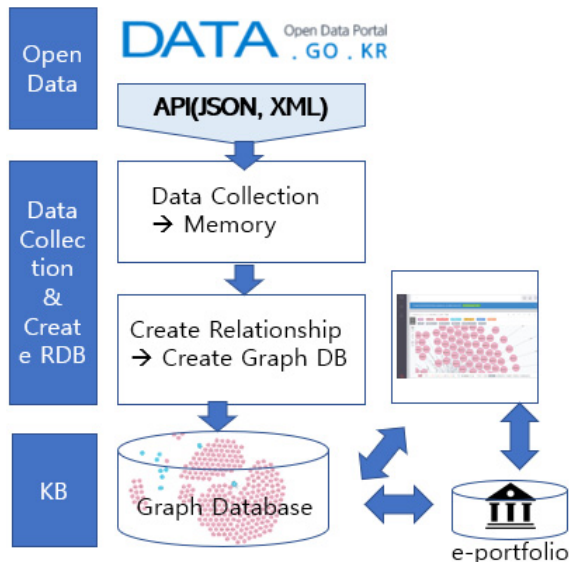


그림 2. 진로 조언 연구 시스템 구조

Fig. 2 Career Advice Research System Structure.

나. 연구 시스템 구축 환경

연구의 효율적인 연구 수행을 위하여 연구 수행 시스템은 표 1과 같이 Microsoft의 클라우드 서비스인 Azure 환경으로 구축하였으며, 데이터 수집 및 처리를 위한 서비스는 Azure App Service를 그래프 기반 지식베이스는 Azure Container 서비스로 시스템 환경을 구

성하였다. 주기적인 데이터 수집 및 데이터 갱신 작업 수행은 Azure WebJobs 서비스를 활용하였다.

표 1. 연구 시스템 구축 환경

Table1. Research system construction environment.

Resource	Environment
App Service	- Stack : .NET - .NET Framework version : V.47
Container Instances	- Properties : 2 CPU cores, 3 Gib Memory - OS Type : Linux - Container image : neo4j:latest
WebJobs	- Run Background tasks

2. 직무데이터 사전 API

직무데이터 사전은 한국고용정보원에서 제공하는 공공데이터로 취업 및 고용을 위한 직무 수행 정보를 탐색할 수 있는 데이터로 관심 직무에 대한 자격증, 학과, 훈련, 주요 용어 등 정보 사전 데이터로 구성되어 있다. 직무데이터 사전은 표 2와 같은 직무 능력에 대한 능력단위, 수행태도, 지식 및 기술, 자격증, 학과, 훈련과정 등의 데이터를 출력한다.

표 2. 직무데이터사전 API 출력 결과

Table2. Job Dictionary API Output Result.

Items	Description
<upperClsCd>	Ability Unit Parent Code
<abltcd>	Ability unit code
<abltCn>	Ability unit code name
<abltDfnCn>	Ability unit definition
<abltUnitLvl>	Ability level
<JobCont>	Job Description
<KnwSkl>	Knowledge and Technology
<att>	Attitude
<crfltem>	Related certifications
<major>	Related Major
<tmg>	Related Training

3. 그래프데이터베이스 설계

현실 세계의 객체들은 서로 간의 관계들이 형성되어 있지만 RDB에 저장되어 있거나 문서화된 데이터들은 대부분 이러한 관계요소를 저장하는데 적합하지 않다. 때문에 이러한 데이터들의 관계 요소를 분석하고 그래프데이터베이스에 적합한 노드와 관계를 구성하기 위해서는 그래프데이터베이스에 적합한 형태로 노드와 관계를 분석·설계해야 한다.

가. 그래프데이터베이스 NODE 설계

그래프데이터는 데이터 요소인 노드, 두 노드를 연결하는 관계인 엣지(Edge)로 구성되며, 속성에 해당하는 프로퍼티를 가질 수 있다. API로부터 수집된 데이터의 형태는 그래프데이터베이스 구조와 큰 차이가 있기 때문에 그래프데이터베이스에 맞게 구조 설계가 이루어져야 한다. 표 3은 직무데이터를 분석하여 설계한 노드와 프로퍼티를 구조이다.

표 3. 노드와 프로퍼티 설계

Table3. design of Node and property.

<pre>(:Data { NcsCd: "", NcsCn: "", NcsCdLvl: "", NcsCdLvlNm: "", Category: "", DtyDfnCn: "" }) (:Ablt { AbltCd: "", AbltCn: "", AbltDfnCn: "", AbltUnitLvl: "" }) (:AbltJobCont { JobCont: "" }) (:AbltKwnSkl { Name: "" }) (:AbltAtt { Name: "" }) (:AbltCrftItem { Name: "" }) (:AbltMajor { Name: "" }) (:AbltTrng { Name: "" })</pre>
--

나. 그래프데이터베이스 데이터 관계 설계

수집된 직무데이터는 표 4와 같이 관계가 설정되지 않은 형태이기 때문에 수집된 데이터의 세부 직무 항목을 분석하여 가장 핵심이 되는 관계 모형을 설계해야 한다.

표 4. API 수집 데이터

Table4. API collection data.

<pre>"jobContList":[{"jobCont":"협력대상국 개발환경 분석하기"}, "knwSkl":"공적개발원조, 국제기구, 독해능력", "att":"감 성적, 계량적, 계획적" "crftItem":"오다(ODA)전문가 3급", "major":"개발 등, 경제, 사회, 정치, 행정", "trngList":null}</pre>

한 개의 직무 항목 데이터 노드는 여러 개의 관계를 가질 수 있기 때문에 수집된 데이터를 분석하여 특정 직무 항목을 하위 노드로 가지는 전체 상위노드를 찾아 관계에 해당하는 엣지를 연결해 주어야 한다. 예를 들어 수집된 직무데이터사전에서 “현실적”이라는 수행태도를 요구하는 능력단위 노드는 228개가 있다.

직무데이터의 경우 표 5와 같이 대/중/소/세분류의 구분 항목은 상위노드들과 ‘[:PARENT_OF]->’의 관계를 의 능력단위(ablt)의 경우 능력을 위한 세부 직무 내용, 지식 및 기술, 수행태도, 자격증, 학과, 훈련 항목과 능력단위 ‘[:항목 구분]->’ 세부항목으로 연결되도록 설정하였다.

표 5. NODE들의 관계 설계

Table5. Node's Relationship Design.

Items	Relationship Design
Category Relationship Design	<pre>(daeBoonRyu:Data)-[:PARENT_OF]->(joongBoonRyu:Data) (joongBoonRyu:Data)-[:PARENT_OF]->(soBoonRyu:Data) (soBoonRyu:Data)-[:PARENT_OF]->(sehBoonRyu:Data) (sehBoonRyu:Data)-[:PARENT_OF]->(ablt:Ablt)</pre>
Detailed items relationship design	<pre>(ablt:Ablt)-[:JOBCONT]->(abltJobCont:AbltJobCont) (ablt:Ablt)-[:KNWSKL]->(abltKwnSkl:AbltKwnSkl) (ablt:Ablt)-[:ATT]->(abltAtt:AbltAtt) (ablt:Ablt)-[:CRFTITEM]->(abltCrftItem:AbltCrftItem) (ablt:Ablt)-[:MAJOR]->(abltJobCont:AbltJobCont) (ablt:Ablt)-[:TRNG]->(abltTrng:AbltTrng)</pre>

3. 그래프 기반 지식베이스 생성

가. Graph database 지식 정보 저장

OpenAPI로 수집된 직무데이터의 관계 설정을 위해서는 표 6과 같이 노드 및 관계 구성을 위한 데이터 전처리 작업을 수행해야 한다. 수집된 직무데이터사전 데이터를 분석·설계하여 연관 관계를 형성하고 그래프데이터베이스에 저장한다. 초기에는 중복 방지를 위해 데이터 생성과 관계형성을 위해 Match를 사용했으나 처리할 때마다 해당 노드의 존재 여부를 확인하기 그래프데이터베이스를 생성하는 데 많은 시간이 소요되었다. 빠른 데이터베이스 생성을 위해 전체 데이터의 중복을 제거와 관련된 노드들의 연관 관계를 설정하여 리스트에 저장하고 이를 각각의 행으로 확장하는 Unwind를 사용하여 80% 정도의 지식베이스 생성속도 향상이 있었다.

표 6. 카테고리 및 세무항목 관계 생성

Table6. Create of Relationship category & Detailed items.

Relationship category :
<pre>.Match("(joongBoonRyuDataNode:Data)", "(soBoonRyuDataNode:Data)") .Create("(joongBoonRyuDataNode)-[:PARENT_OF]->(soBoonRyuDataNode)")</pre>
Relationship Detailed items :
<pre>.Match("(abltNode:Ablt)", "(abltAttNode:AbltAtt)") .Where((GraphDB.Ablt abltNode) => abltNode.AbltCd == ablt.AbltCd) .Create("(abltNode)-[:ATT]->(abltAttNode)")</pre>

나. 그래프데이터베이스 NODE 사이즈와 구조

그림 4는 수집된 직무데이터에 관계요소를 설계 및 전처리하여 최종적으로 그래프데이터에 입력한 형태이다. 직무데이터 사전 내의 개체에 대입되는 노드는 총 56,034개가 생성되었고 최종 그래프데이터베이스에 저장된 노드 항목, 수량, 연결의 형태는 그림 3와 같다.

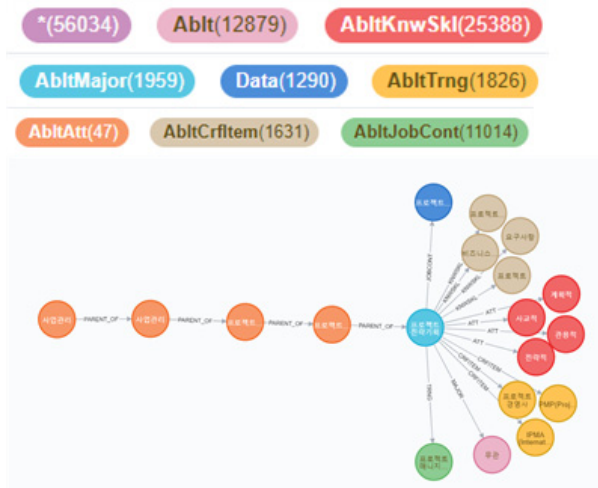


그림 3. 저장된 그래프 구조

Fig. 3. Stored Node Size and Grid DB Structure.

다. Graph database Edge 사이즈

관계가 설정되지 않은 직업사전 공공데이터에 실제 관계를 설정하여 생성된 엣지의 수는 표 7과 같이 총 342,758개이다. 연구의 효율화를 위해 단방향의 엣지를 생성하였으나 실제 노드 서로간의 양방향 관계, 자신과의 관계와 같은 다양한 관계 요소들이 추가될 수 있다. 그래프데이터베이스에서는 이러한 엣지의 관계가 가장 중요한 분석 요소이기 때문에 데이터가 증가하고 다양해질수록 관계에 대한 설계와 관리가 중요하다.

표 7. 저장된 Edgn 사이즈 및 합계

Table7. Stored Edge Size & Amount.

Items	Description	Amount
ATT	Attitude	134866
CRFITEM	Certificate	39,468
MAJOR	University Major	21,943
KNWSKL	Knowledge and Skill	128,447
PARENT_OF	Parent node	1,290
JOBCONT	Job Description	23,893
TOTAL		342,758

IV. 그래프데이터베이스 능력 조언 모델 연구

1. 학생 맞춤형 직무 능력 추천

e-포트폴리오를 활용하는 대학마다 제공하는 기능 및 항목의 차이가 크다. 연구를 위해 가상의 e-포트폴리오에 저장된 데이터를 기준으로 컴퓨터공학 전공, 기술적 수행태도와 정보보안기사 자격증, 정보화 사업 능력과 사전점검 능력을 보유하고 있는 학생에 적합한 직무 능력을 표 8과 같이 관계를 사이퍼 쿼리로 질의했을 때 ‘보안성 검토’ 직무능력으로 답변되었다. 답변된 ‘보안성 검토’의 지식맵은 그림 4와 같이 연관된 노드와의 관계를 시각화하여 확인할 수 있으며, 질의된 능력 단위와의 연결된 경로를 검증이 가능하다.

표 8. 맞춤형 능력 추천 사이퍼쿼리

Table8. Customized ability recommendation cypher query.

```

MATCH
(n:Abtl)-[:ATT]->(att),(n)-[:MAJOR]->(major),(n)-[:CRFITEM]->(crfitem), (n)-[:KNWSKL]->(knwskl)
WHERE (n)-[:ATT]->({Name:'기술직'})
AND (n)-[:MAJOR]->({Name:'컴퓨터공학'})
AND (n)-[:CRFITEM]->({Name:'정보보안기사'})
AND (n)-[:KNWSKL]->({Name:'정보화사업'})
AND (n)-[:KNWSKL]->({Name:'사전점검'})
RETURN n
  
```

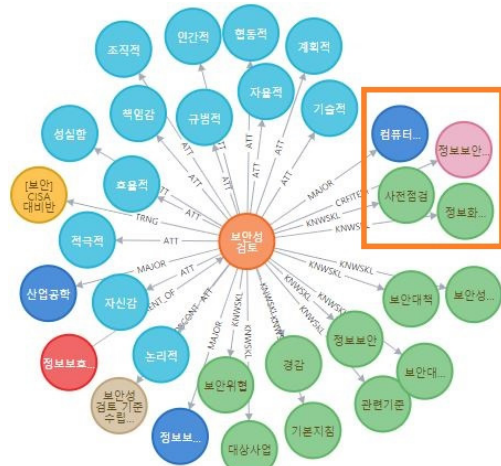


그림 4. 보안성 검토 능력 지식맵(expand relationships)

Fig. 4. Security Review Ability Knowledge Map.

2. 대화형 인터페이스를 환경 직무능력 조언

조언 시스템의 활용성 테스트를 위하여 대학생들에게 가장 친숙한 인터페이스인 대화형 인터페이스를 활용하여 학생 맞춤형 직무능력으로 추천받은 ‘보안성 검토’ 능력을 갖추기 위한 기술, 자격증을 질문하였다.

대화형 인터페이스 구현을 위한 챗봇 프레임워크로는 Microsoft BotFramework V4를 활용하였으며, 사용자의 질문에 대한 자연어 처리 부분은 Microsoft의 NLP (Natural Language Processing) 서비스인 Luis를 활용하여 진로조언을 위한 자연어 학습을 진행하였다.

사용자의 자연어 질문은 Luis 서비스를 통해 표 9와 같이 사용자의 Intent와 Entity를 전달받아 '보안성 검토'에 대한 자격증과 지식 및 기술 관계를 정의하였으며 그림 5과 같이 사용자의 질문 의도를 분석하여 그래프 기반 지식베이스를 활용한 직무 능력과 관계 내용을 답변할 수 있었다.

표 9. 자연어 처리와 직무 능력 관계 질의
Table9. NLP and competency relationship query.

Recognizer Result	Raw Response
<pre>{ "recognizerResult": { "alteredText": null, "entities": { "\$instance": { "세부능력": [{ "endIndex": 14, "startIndex": 11, "text": "자 격 증", "type": "세부능력" }, { "endIndex": 18, "startIndex": 16, "text": "기 술", "type": "세부능력" }] } } } }</pre>	Top-Scoring Intent 직업 (0.9063048) Please add your LUIS service to enable reassigning. Entities 세부능력 → 자격증, 지식및기술 직업 → [\"직업리스트\", \"보안성 검토\"]

자격증 정보:
query.Match("(competency)-[RelatedCertification]->(Certifications)")

지식 및 기술 정보:
query.Match("(competency)-[Related Knowledge]->(Knowledge)")

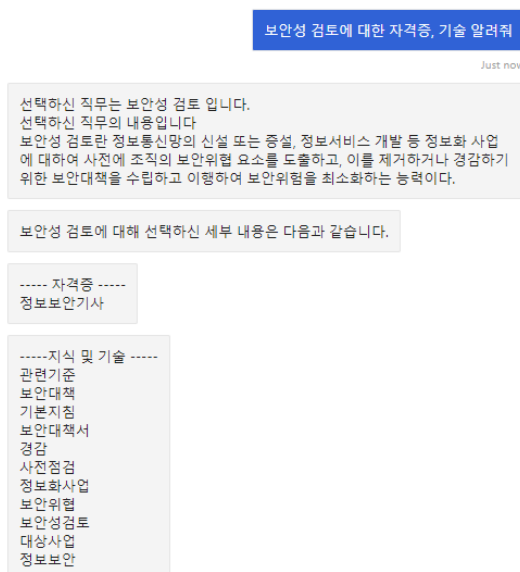


그림 5. 대화형 인터페이스 활용 조언 모델 테스트
Fig. 5. Interactive model utilization advice model test.

3. 그래프 기반 지식베이스 활용 진로 조언 모델 연구 결과

대학생 진로 조언을 위한 공공데이터 활용 그래프 기반 지식베이스 구축과 활용성 연구를 위해 관계가 설정되지 않은 공공데이터를 수집하여 연관 관계를 분석·설계하고 그래프 기반 지식베이스 생성과 생성된 지식베이스를 가상의 대학생 e-포트폴리오 정보와 매핑하여 학생 개인에 맞춤형 능력 추천과 조언이 가능한 대화형 인터페이스 환경을 구축하였다.

구축된 시스템은 사용자 데이터의 부족과 수집할 수 있는 공공데이터의 한계로 인하여 복잡한 관계를 분석과 같은 성능 및 효율 분석에는 한계가 있지만, 공공데이터를 활용한 연관 관계 구성과 그래프데이터베이스 기반의 지식베이스를 구축 및 이를 활용한 조언 시스템 모델 구현을 통해 개인화된 데이터와의 매핑을 통한 추천과 조언 시스템으로서의 활용 가능성을 검증하였다.

V. 결 론

학생 맞춤형 진로 조언은 학술적 연구와 인력 양성의 목적을 가진 대학의 목적을 달성을 위한 매우 중요한 요소이다. 그러나 대학생들의 가장 큰 고민인 진로문제 해결을 위한 정보기술 활용에 대한 연구는 아직 미흡한 것이 사실이다. 본 연구를 통해 e-포트폴리오와 같은 대학생의 개별 데이터를 그래프 기반 지식베이스에 매핑하여 자신에게 최적화된 직무와 역량을 추천받고, 이를 개발하기 위한 커리큘럼이나 수행태도, 자격증 등의 진로 정보를 지식맵을 통해 확인할 수 있으며, 대학생들은 자신이 진로 목표 달성을 위한 현재 상태와 부족한 부분을 확인하여 보완하는 데 도움이 될 수 있다.

현대 시대의 데이터들은 데이터 상호 간에 다양한 연관 관계를 맺고 있는 경우가 많지만 대부분의 정보시스템은 이러한 연관 관계가 무시된 채 활용되고 있다. 이러한 관계 속성에 대한 활용 연구를 위해 한국고용정보원의 직무데이터 사전에 API를 통해 수집 및 분석하여 연관 관계 설계와 구성을 통한 그래프데이터베이스 기반의 지식베이스를 구축하였다. 구축된 지식베이스에 가상의 학생 e-포트폴리오 데이터를 매핑하여 맞춤형 직무 추천과 대화형 인터페이스 환경에서의 조언 시나리오 검증을 통한 진로조언 시스템 구축과 활용 방향을 제시하였다.

그러나 과거와 달리 현대의 직업정보는 기술과 환경의 급격한 변화로 빠르게 변화되고 있기 때문에 오픈데

이터만으로는 학생들에게 진로 조언을 할 수 있는 지식 베이스 구축에 한계가 있다. 때문에 실제 활용을 위한 데이터 수집을 위해서는 실제 기업들의 최신 구인 정보를 신속하게 수집하여 직무데이터와 속성데이터 설계 및 분류를 통한 지식베이스 구축과 수집된 데이터의 연관 관계 활용에 대한 추가 연구가 필요하다.

REFERENCES

- [1] HyunJin Jang, Survey of College Career Education Status(2017) -Basic Statistics Table-, Ministry of Education, pp. 39-304, 2017.
- [2] Bandura, A. and Simon, K. M., The Role of Proximal Intentions in Self-regulations of Refractory Behavior, Cognitive Therapy and Research, Vol. 1, No. 3, pp. 177-193. September 1977.
- [3] JongChan Lee, A comparative analysis of theoretical considerations and research results on career decision self-efficacy of college students, Career research, Vol. 3, No. 1, pp.61-62, June 2013.
- [4] Sunjong Choi, Studies on prospects and directions of the policy for young people, Korea Youth Research Association, Vol. 26, No. 1, pp. 253-273, January 2019.
- [5] Yeon Soon Jung, A Study on Young Adults' Job Selection - The Case of Civil Service Examinees, Journal of Vocational Education & Training, Vol. 13, No. 3, pp. 123-146, December 2010.
- [6] L. C. Page and H. Gehlbach., How an artificially intelligent virtual assistant helps students navigate the road to college. AERA Open, Vol. 3, no. 4, pp. 8-10 December 2017.
- [7] Jong-wook Kim, Why do you want to do your first job so quickly?(Monthly labor diary), KOREA LABOR INSTITUTE, pp. 22-35, 2017.
- [8] <http://www.sba.seoul.kr/kr/sbbs01s1/5033669>
- [9] Jong-yul Kim and Kwang-hyun Roh, A Study on the Data Collection and Convergence of Career Advisor System Using AI, Journal of Digital Convergence, vol.17, no.2, pp. 177-185, February 2019.
- [10] B. Williamson, The hidden architecture of higher education: building a big data infrastructure for the 'smarter university', International Journal of Educational Technology in Higher Education, Vol. 15, no.1, pp. 12, March 2018.
- [11] Michael, T. and Hillmann, S. and Weiss, B., Alex: An Artifical conversational agent for students at the tu berlin, 28. Konferenz Elektronische Sprachsignalverarbeitung, pp. 238-245, 2017.
- [12] Ae-Kyung Choi, A Study on Establishing and Managing an e-Portfolio System for Career Development of College Students, The Journal of Business Education, vol.13, No.8, pp. 169-192, August 2006.
- [13] Minsu Jang and Jaehong Kim, Social Robots in the Fourth Industrial Revolution, The Magazine of the IEIE, vol.45, No.9, pp. 35-43, September 2018.
- [14] H. Jeong, Structure and Application of Complex Network, The Magazine of the IEIE, vol. 31, No.4, pp. 438-444, April 2004.
- [15] Ian Robinson and Jim Webber and Emil Eifrem, Graph Databases, 2nd Edition, O'Reilly Media, pp 2-24, 2015.
- [16] Leyla Zhuhadar and Mark Ciampa, Leveraging learning innovations in cognitive computing with massive data sets: Using the offshore Panama papers leak to discover patterns, Computers in Human Behavior, 2017.

저 자 소 개



김 종 울(학생회원)
2016년 방송통신대학교대학원 이
러닝학과 이학석사
2015년~현재 트리포드(주) 컨설
팅본부 부장
2019년 한성대학교 대학원 스마트
융합컨설팅학과 박사과정

<주관심분야: 클라우드컴퓨팅, 에듀테크, 빅데이터
분석>



노 광 현(정회원)
1995년 고려대학교 산업공학과
공학사
1997년 고려대학교 산업공학과
공학석사
1997년 고려대학교 산업공학과
공학박사

2002년 Ecole des Mines de Paris, Post-Doc.
2003년~2006년 한국전자통신연구원 연구원
2006년~2007년 한국항공우주연구원 선임연구원
2007년~현재 한성대학교 IT융합공학부 교수
<주관심분야: IoT, 지능형시스템>