

## 기업 직무 정보를 활용한 OOPP(Optimized Online Portfolio Platform)설계

정보근\*, 박진욱\*\*, 이병관\*\*\*

### A Design of the OOPP(Optimized Online Portfolio Platform) using Enterprise Competency Information

Bogeun Jung\*, Jinuk Park\*\*, ByungKwan Lee\*\*\*

**요약** 본 논문에서는 직무별로 취업에 필요한 역량을 나타내고, 구직자가 온라인상에서 포트폴리오를 효율적으로 작성하고 관리하는 OOPP(Optimized Online Portfolio Platform)를 제안한다. 제안하는 OOPP는 세 가지 모듈로 구성된다. 첫째, JDCM(Job Data Collection Module)은 직업정보 사이트의 구인 광고들을 수집하여 스프레드시트에 저장한다. 둘째, CSM(Competency Statistical Model)은 수집한 구인 광고들을 텍스트 마이닝하여 직무별로 요구되는 핵심 역량을 분류한다. 셋째, OBBM(Optimize Browser Behavior Module)은 브라우저의 처리속도를 개선하여 사용자가 데이터를 빠르게 조회할 수 있게 한다. OBBM은 검색엔진의 연산을 최적화하는 PSES(Parallel Search Engine Sub-Module)과 이미지 텍스트 등의 로드를 최적화하는 OILS(Optimized Image Loading Sub-Module)로 구성된다. 제안하는 OOPP의 성능분석 결과 CSM로 분석된 데이터의 정확도는 최대 100%, 최소 99.4%로 실제 광고와 분석된 데이터의 차이가 거의 발생하지 않았으며, OBBM을 이용한 브라우저 최적화를 실행하면, 작업시간이 약 68.37%가 감소한다. 결과적으로 OOPP는 직현재 직업정보 사이트의 구인 광고를 정확하게 분석하여 사용자가 분석한 결과를 웹페이지에서 신속하게 조회할 수 있다

**Abstract** This paper proposes the OOPP(Optimized Online Portfolio Platform) design for the job seekers to search for the job competency necessary for employment and to write and manage portfolio online efficiently. The OOPP consists of three modules. First, JDCM(Job Data Collection Module) stores the help-wanted advertisements of job information sites in a spreadsheet. Second, CSM(Competency Statistical Model) classifies core competencies for each job by text-mining the collected help-wanted ads. Third, OBBM(Optimize Browser Behavior Module) makes users to look up data rapidly by improving the processing speed of a browser. In addition, The OBBM consists of the PSES(Parallel Search Engine Sub-Module) optimizing the computation of a Search Engine and the OILS(Optimized Image Loading Sub-Module) optimizing the loading of image text, etc. The performance analysis of the CSM shows that there is little difference in accuracy between the CSM and the actual advertisement because its data accuracy is 99.4~100%. If Browser optimization is done by using the OBBM, working time is reduced by about 68.37%. Therefore, the OOPP makes users look up the analyzed result in the web page rapidly by analyzing the help-wanted ads. of job information sites accurately.

**Key Words** : OOPP(Optimized Online Portfolio Platform), Optimized Browser, Text mining, competency, help-wanted

\*Computer Engineering of Graduate School, Catholic Kwandong University

\*\*Corresponding Author : Computer Engineering of Graduate School, Catholic Kwandong University

\*\*\*Computer Engineering of Graduate School, Catholic Kwandong University

Received September 20, 2018

Revised September 30, 2018

Accepted October 01, 2018

## 1. 서론

현재 취업 시장은 스펙 중심의 채용보다는 직무수행 역량 중심의 채용으로 변화하고 있다. 갈수록 상향 평준화되는 구직자들의 스펙 때문에 기업들은 일정 수준 이상의 스펙을 입사조건으로 설정하기도 한다. 현재 구직자들은 자신의 직무수행 능력을 증명하기 위하여 스펙 뿐만 아니라 포트폴리오를 작성하여 기업에 제출하고 있다. 기업은 구직자의 포트폴리오를 통하여 구직자의 직무 역량을 판단하고, SNS를 통하여 구직자와 활발히 소통한다. 기업의 채용담당자들은 구직자의 SNS활동을 조회하거나, 본인이 직접 구인 SNS페이지를 운영하여 실시간으로 구직자의 의문점들을 해소해주고 있다.

이런 상황에서 구직자의 SNS활동은 구직자의 또 다른 포트폴리오가 될 수 있다. 실제로 마이크로소프트에서 운영하는 웹페이지인 'Linkedin'은 구인/구직자들이 자신이나 기업의 정보를 공유하는데 사용하는 SNS다 [1]. 그러나 'Linkedin'을 제외한 국내 직업정보사이트는 여러 기업의 구인 광고들을 웹에 게시하고, 구직자는 제한된 형식의 이력서를 작성하는 방식으로 운영되고 있다. 따라서 본 논문에서는 구직자가 지원하려 하는 직무의 핵심 역량을 쉽게 조회하고 자신만의 포트폴리오를 작성할 수 있는 OOPP(Optimized Online Portfolio Platform)를 제안한다.

OOPP는 직업정보사이트에서 정보를 수집하는 JDCM(Job Data Collection Module), 수집한 정보를 정규화하여 유의미한 데이터로 만든 후 직무별 핵심 역량을 분석하는 CSM(Competency Statistical Model)과 웹페이지의 연산 방식과 이미지, 텍스트 등 리소스의 로드를 최적화하는 OBBM(Optimize Browser Behavior Module)으로 이루어진다. 본 논문은 다음과 같이 구성된다. 본 논문의 2절은 OOPP에 사용되는 기술들의 관련 연구들을 설명하고, 3절은 OOPP의 기술적인 설계방법과 동작 과정을 설명한다. 4절은 OOPP의 유효성을 검증하기 위하여 CSM의 정확성과 OBBM의 효율성에 대하여 실험을 진행한다. 마지막으로 5절은 본 논문의 종합적인 결과를 설명한다.

## 2. 관련 연구

### 2.1 포트폴리오

대학교육에서 기존 교수학습 운영 시스템의 한계를 극복할 수 있는 대안으로써, 또한 교수의 교수 활동과 학생 학습활동의 결과뿐 아니라 성장 과정을 스스로 관리, 개발해 나갈 수 있는 시스템으로써 'e-교수학습 포트폴리오'를 제시하고, 이를 구현하기 위한 Framework와 그 구성요소를 개발하고, 활용 방법으로 이용했다. e-교수학습 포트폴리오 Framework를 개발영역과 개발원리에 따라, 교수영역, 수업영역, 학습영역과 성장, 성찰, 교류, 관리로 구분하고, 주요 구성요소를 도출했다. [2]

기존의 e-포트폴리오 관리시스템의 문제점을 분석하고 이를 극복할 수 있는 새로운 e-포트폴리오 관리 시스템은 다른 사람들과 경쟁하는 구조로 되어 있다. 모든 학생의 활동은 블로그와 같은 환경에서 추천과 상호 참여를 유발하도록 이용했다. [3]

전자포트폴리오는 컴퓨터 기술을 활용하여 문서 중심 포트폴리오 평가의 현실적 문제점들을 극복하고 수업의 목적, 내용과 방법, 그리고 평가를 수행중심으로 통합시킬 수 있는 강력한 도구가 될 수 있다. 웹 기반 전자포트폴리오의 개발과 적용을 통해 전자포트폴리오가 수행평가의 적용을 보다 활성화할 수 있을 뿐 아니라, 학습자들의 자발적이고 적극적 학습을 촉진할 수 있다. [4]

RBC제도가 요구하는 조건에서 보험회사의 평균-분산 효율을 최대화하는 포트폴리오를 찾는 문제를 수학적으로 정의하고 이 문제의 최적해를 구하는 방법을 제시한다. 제시된 최적화 방법은 위험회피계수, RBC 비율, 보험회사의 자본 규모 등의 주요 변수들의 변화가 보험회사의 최적 포트폴리오와 보험회사의 수익률에 미치는 영향을 연구하는데 활용된다. [5]

대학 환경에 적합한 e-포트폴리오 구현모델인 K-folio의 시스템 아키텍처는 대학교육을 위한 e-포트폴리오 이론적 Framework에 근거하여 설계했다. 개발 방법론으로는 신속응용시스템개발방법(RAD)을 채택하였는데, 이는 기본적인 핵심기능을 구현한 후 시범운영을 통하여 사용자 요구사항을 체계적으로 반

영하면서 시스템의 완성도를 높였다. [6]

유치원교사를 평가하기 위한 포트폴리오를 객관적으로 점수화할 수 있는 평가 루브릭은 선행연구와 전문가 및 유치원교사의 내용 타당도 검증에 근거하여 1차 루브릭을 구성했다. 다음으로 10인의 유치원교사들이 제작한 포트폴리오를 구성된 루브릭으로 평가하게 한 후, 이에 대한 피드백을 거쳐서 개발했다. [7]

선물 간의 의존도(dependence)를 고려하여 포트폴리오의 위험을 측정함으로써 초기증거금의 설정에 대해 알아보기 위해서 7가지 농업 원자재 선물로 이루어진 포트폴리오의 신용위험을 Value-at-Risk, Expected shortfall, Spectral risk measure와 같은 리스크 측도를 사용하여, 선물 간의 의존성을 고려할 때 리스크 측정치가 어떻게 변하는지 확인했다. 선물 간의 의존성은 코플라로 반영하고, 각 선물의 손실 분포는 normal inverse Gaussian을 이용하여 모형화했다. [8]

생태계 기반 어업관리의 적용기법 마련을 위한 기초적 연구로 포트폴리오 기법을 이용하여 복수어종의 최적 생산 수준을 실증적으로 분석했다. 실증분석은 복수어종의 대형 선망어업을 대상으로 했다. 대부분 어업은 복수어종을 어획 대상으로 하고 있고, 복수어종 어획에 있어 특정 어종에 대한 어획은 생태적 혹은 기술적으로 연관된 다른 어종의 어획과 경제적 가치(수익)에 영향을 미치게 된다. 이에 따라 포트폴리오 기법을 적용하여 어종 간 생물학적 그리고 기술적 상호연관성에 따른 기대수익률과 위험 분석을 통해 복수어종 내 개별 어종의 최적 생산계획을 수립할 수 있다. [9]

## 2.2 Rendering

대화형 프레임 속도에서 3D 라인 데이터의 렌더링을 위한 기술을 이용한다. 촘촘한 라인 번들을 강조하기 위해 선 주위에 깊이 의존적인 후광을 작성하는 반면 덜 구조화 된 선은 강조하지 않는다. 또한, 선평 감쇄를 통한 깊이 큐와 결합 된 깊이 종속 후광은 깊이 인식을 증가시켜 밀도가 낮은 선 렌더링에서 고밀도 선 데이터의 시각화까지 확장하여 응용한다. [10]

고성능 카툰 렌더링은 일반 카툰 렌더링에서 표현

하기 힘든 질감을 만화적인 스타일로 표현한다. 만화적인 질감을 표현하기 위해 [11]은 GPU 프로그래밍을 이용한 노말 매핑 방법을 사용하였다. 해당 기술의 방법은 물체의 복잡도 증가 없이 단순한 물체에 만화적인 질감을 표현할 수 있어 창조적인 디지털 콘텐츠 분야나 게임 시스템에 활용될 수 있다.

광 회절 단층 촬영 (Optical Diffraction Tomography, ODT) 기법은 세포 영상을 3차원 가상 공간에 시각적으로 표현하고 기존의 세포 영상들과의 일치감을 주는 색상 매핑 기술을 포함한 가시화 프레임워크를 사용한다. 3차원 형상을 실시간으로 렌더링할 수 있는 가시화 기법을 응용한다. 해당 기법에서는 우선 세포의 3차원 형상을 나타내기 위해 볼륨 데이터의 가시화에서 널리 활용되고 있는 볼륨 렌더링 기법을 ODT 영상에 맞게 적용했으며, 빈 공간 교란 기법을 통한 렌더링 결과의 개선으로 세포내 구조의 연속성을 나타낼 수 있게 했다. [12]

연필 드로잉의 느낌을 주는 영상으로 변환하는 방법을 응용했다. 이 방법은 선회선 적분 (LIC, Line Integral Convolution)이라는 기법을 이용해서 구현했다. 먼저, 영상으로부터 널리 알려진 코히런트 라인 (coherent line) 기법을 적용해서 특징선을 추출한 다음, 영상을 배경, 특징선 주변, 특징선으로 구분하여 각각에 특성에 맞는 연필 렌더링 영상을 LIC를 적용해서 생성한 다음, 이를 적절하게 합성함으로써 게임 캐릭터에 대한 연필 드로잉 기법을 구현한다. [13]

## 2.3 HTML

HTML5 canvas를 활용하여 시각화 도구와 사용자와의 동적 상호작용을 지원하는 웹 기반의 시각적 공간분석 환경을 설계하여 사용했다. 분석 환경은 군집 지도, 애니메이션 지도, 시간적 평행좌표 그림, 시계열 히트 맵 차트와 같은 시각화 도구로 구성되어 있다. 분석 환경에서는 동일한 군집으로 분류된 지역의 속성 변화를 살펴볼 수 있는 다중창 배열과 선별적 강조를 기반으로 하는 패턴 탐색이 가능하다. 그리고 크로스 브라우저를 지원하기 때문에 다양한 정보 기기를 사용하는 컴퓨팅 환경에서도 활용할 수 있다. [14]

핵심 통찰력을 응용한 HTML은 프로그램이 원시

HTML 마크업 보다 문서 구조에 동적으로 액세스 할 수 있게 해주는 W3C 지정 인터페이스인 DOM 트리로 작업하는 방법을 이용한다. HTML 웹 페이지에서 콘텐츠를 추출하기 위해 공개적으로 사용 가능한 웹 프록시에서 접근 방식을 이용했다. 이 프로시는 중앙 집중식으로 사용할 수 있으며 개인 그룹뿐만 아니라 개인 그룹에 대해서도 관리 할 수 있다. [15]

NCSLab (Networked Control System Laboratory) 프레임 워크를 기반으로 한 대화식 가상 배선을 위해 3D 모델링, HTML5 기반 렌더링 및 제어 알고리즘 설계를 포함한 방법이 사용된다. 해당 방법은 NCSLab에서 이미 구현 된 제어 알고리즘 설계와 결합되어 사용자가 MATLAB / Simulink 실시간 워크샵을 사용하여 알고리즘을 사용자 정의 할 수 있다. 원격으로 제어 알고리즘을 구현하는 것 외에도 온라인 가상 실험이 제대로 진행되기 전에 3 차원 가상 배선 프로세스가 올바르게 완료되어야 한다. [16]

### 2.4 텍스트 마이닝

텍스트 마이닝은 전자문서로부터 분석이 가능한 유의미한 자료를 끌어내고, 실제로 끌어낸 자료를 기반으로 군집 분석, 주요 키워드 추출 등 유용한 정보를 생성한다.

전자문서를 유의미한 자료로 변환하기 위하여, 한글의 경우 R을 기반으로 KoNLP 패키지를 사용하여 텍스트를 분석한다. KoNLP는 KAIST 품사 태그셋을 사용하여 한글 텍스트를 분석한다[17]. 그림 1은 KoNLP에서 사용하는 한글 품사 태그셋을 나타낸다. KoNLP를 사용하여 텍스트를 분석한 연구들은 다음과 같다.

부경대학교 대학원은 한글 텍스트 마이닝을 이용하여 인터넷 뉴스를 분석한다. 분석에는 RHadoop이 사용되었으며, 모든 과정을 유저인터페이스 환경으로 구축하여 연구의 결과물을 실생활에 사용하고자 하였다. 또한, 반응형 웹을 적용하여 플랫폼의 제약이 없었다 [18].

서울대학교 건설환경공학부는 위치기반 소셜 미디어 데이터의 텍스트 정보를 분석하여 주요한 키워드들이 공간적으로 어떻게 분포하고 있는지를 파악할 수 있는 일련의 분석방법론을 적용하였다[19].

안길승 외 2명은 작업 유형, 원인적 원인 물질, 사고 유형 및 사고 보고서의 사망자 수를 추출하는 사고 분류 모델을 제안했다. 분류 모델은 단어 발생 기능이 있는 SVM (Support Vector Machine)이며 이러한 기능은 상호 정보를 기반으로 선택된다. 분류 모형과 온톨로지는 다양한 사고를 효과적으로 분석한다 [20].

## 3. OOPP(Optimized Online Portfolio Platform)설계

### 3.1 Overview

본 논문에서 제안하는 OOPP(Optimized Online Portfolio Platform)는 직무별로 취업에 필요한 역량을 나타내고, 구직자가 온라인상에서 포트폴리오를 효율적으로 작성하고 관리한다. OOPP는 다음과 같이 세 가지 모듈로 구성된다.

첫째, JDCM(Job Data Collection Module)은 웹 페이지의 xml정보를 스프레드시트에 저장한다.

둘째, CSM(Competency Statistical Module)은 수집한 정보를 정규화하여 직무별 핵심 역량을 분류하

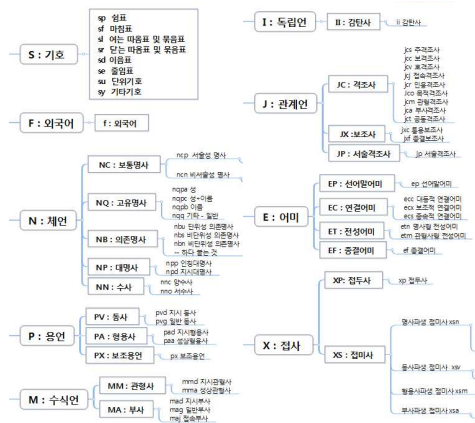


그림 1. 한글 품사별 태그셋  
Fig. 1. Tags set for parts of speech in Hangul

고, 각 직무에서 주로 어떤 역량들을 요구하는지 분석한다.

셋째, OBBM(Optimize Browser Behavior Module)은 브라우저의 로드속도를 최적화하여 사용자에게 신속하게 정보를 제공한다.

OBBM은 병렬 검색엔진 모듈인 PSES (Parallel Search Engine Sub-Module)를 사용하여 브라우저의 처리속도를 향상시키고, OILS(Otimized Image Loading Sub-Module)를 사용하여 이미지, 텍스트 등 웹페이지 로드에서 필요한 자료의 처리를 최적화한다.

### 3.2 JDCM(Job Data Collection Module)

#### 설계

구직자가 원하는 직무의 핵심 역량에 대한 포트폴리오를 효율적으로 작성하기 위해서, 본 논문은 구인 사이트의 데이터를 수집하는 JDCM(Job Data Collection Module)을 설계한다. JDCM은 웹 데이터를 그림 2와 같이 Importxml 기법을 사용하여 구글 스프레드시트에 저장한다. 웹 데이터란 특정 직무에서 요구하는 역량들이다.

Importxml 기법은 구글 스프레드시트에서 웹상의 데이터를 가져와 필요한 데이터를 추출할 때 사용하는 기법이다. Importxml 기법은 Import io기법과 webservicefilterxml 기법과 달리 원하는 정보만을 효율적으로 가져올 수 있다. Importxml 기법은 다른 페이지의 정보를 가져오더라도 페이지 주소의 정보만 바뀌면 다른 작업이 필요 없이 자동 채우기를 통해 정보를 입력할 수 있다는 장점이 있다.

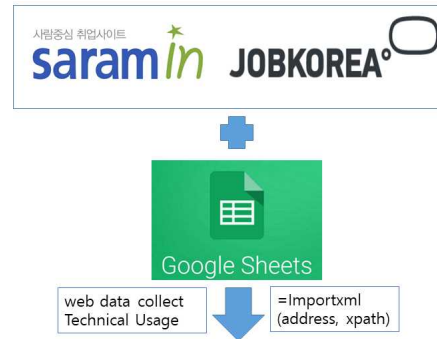


그림 2. JDCM의 구성도  
Fig. 2. the structure of the JDCM

Importxml 기법을 사용하여 직업정보 사이트인 '사람인', '잡코리아' 등의 사이트에서 취업 관련 역량정보를 가져왔다. 두 사이트 모두 같은 방식으로 정보를 수집할 수 있다. Importxml은 URL, Xpath\_query 형식으로 되어 있으며 괄호 안의 왼쪽은 사이트의 주소, 오른쪽에는 필요데이터의 xpath를 넣는다. 필요한 직무에 관련된 역량정보만을 얻기 위하여 사이트의 주소와 수집정보의 xpath 위치를 활용하고 Importxml 기법을 사용해 엑셀로 정보를 갖고 온다. Importxml은 크게 5가지 과정으로 분류한다.

첫째, 구글 스프레드시트를 실행하여 함수 Importxml을 사용한다. Importxml을 사용하면 페이지의 주소와 필요한 정보가 있는 곳의 xpath를 가져온다. xpath는 가져오려는 정보의 위치 정보다. 구글 스프레드시트는 엑셀과 기능은 대부분 비슷하지만, 실시간으로 글이 저장되어 안정성이 높으며 인터넷상에서 공유할 수 있다.

둘째, 직업정보사이트에 접속하여 원하는 정보를 취업 정보를 찾는다.

셋째, 구직 관련 글로 접속한 다음 원하는 정보의 사이트 주소와 xpath 데이터를 얻는다. 예를 들어 그림 3은 직업정보 사이트인 '사람인'의 xpath 데이터 창을 나타낸다. 아래에 파란색 범위에 해당하는 곳이 원하는 정보의 xpath 값이다.

```
<!DOCTYPE html>
<html lang="ko">
<head></head>
<body id="r1_layout_pgsar">
<section id="r1_section"></section>
<script type="text/javascript">var isPersonLogin = false; var...
</script>
<script type="text/javascript">var webBaseurl = ""; var isUrl...
</script>
<div class="iframe_layer" id="iframe_layer" style="display: none;">
</div>
<div id="pop_login_layer_dialog" style="background: url(//www.saraminimage.co.kr/event_2014/0708_rank1/bg_dialog.png) 0px 0px; left: 0px; top: 0px; width: 100%; height: 100%; display: none; position: fixed; z-index: 9999;"></div>
<div class="pop_login_layer" id="pop_login_layer" style="top: 280px; display: none; position: absolute;"></div>
<script type="text/javascript">(function() { var psProto =...
</script>
<script src="//www.saraminimage.co.kr/js/uid6.js" type="text/javascript"></script>
<script type="text/javascript">_u._sid = "saramin"; _u._uid_con...
</script>
<script>try { if(document.com...</script>
<script type="text/javascript">function _u_trackEvent(category...
</script>
<script src="//nxs.naver.net/uclog.js" type="text/javascript">
</script>
<script type="text/javascript">if (ucs_add) var ucs_add(); -
</script>
<script type="text/javascript">window.NREUM||(NREUM={});NREUM...
</script>
</body>
</html>
```

그림 3. xpath 데이터 출력  
Fig. 3. xpath data output

넷째, 그림 4와 같이 스프레드시트에 필요한 주소의 코드를 입력하고 자동 채우기 기능을 사용하면 그림 5와 같이 정보가 입력된다.

이와 같은 방법으로 주소 코드만 변경한 후에 자동 채우기 기능을 사용하며 많은 양의 데이터를 효율적으로 수집하고 사용할 수 있다.

주소코드	A	B	C	D	E	F
005930	조대훈(2.3년제)	Java 기반 시스템 소프트웨어개발 관련 경력				
068270	JAVA 능숙자	DevON 프로젝트 운영 업무 처리 가능자				
005562	융합시스템 UI 설 UI 개발 경험자	연계시스템 설계 테스트 설계 경험 C#사용 사용자				
007356	전기전자통신기 영어 가능자	관련 분야 자격증 각 직무별 Tool 사용 가능자				
095572						

그림 4. 자동 채우기 실행  
Fig. 4. auto-fill execution process

주소코드	A	B	C	D	E	F
005930	조대훈(2.3년제)	Java 기반 시스템 소프트웨어개발 관련 경력				
068270	JAVA 능숙자	DevON 프로젝트 운영 업무 처리 가능자				
005562	융합시스템 UI 설 UI 개발 경험자	연계시스템 설계 테스트 설계 경험 C#사용 사용자				
007356	전기전자통신기 영어 가능자	관련 분야 자격증 각 직무별 Tool 사용 가능자				
095572	스타트업 경험자 온라인 서비스 및 서비스마케팅 지 통계패키지 사용 데이터 분석 사용					

그림 5. 자동 채우기 완성  
Fig. 5. auto-fill completion

### 3.3 CSM(Competency Statistical Module)

#### 설계

본 논문에서 제안하는 CSM(Competency Statistical Module)은 직업을 구하는 사람들을 위하여 현재 회사들이 직무별로 선호하는 역량이 무엇인지 분류한다. CSM은 특정 직무를 목표로 직업을 구하는 사람들이 더 쉽게 핵심역량을 개발하고, 역량에 대한 포트폴리오를 작성할 수 있게 한다.

CSM은 텍스트 마이닝을 사용한다. CSM은 회사가 필요한 인재의 역량들을 입력으로 사용하여, 구직에 필요한 직무별 역량이 무엇인지 분석한다.

우선 CSM은 여러 회사가 인터넷상에 등록된 구인 광고들을 수집한다. 다음으로 CSM은 수집한 광고들에 전처리과정을 적용하여 유의미한 데이터로 정규화한다. 마지막으로 CSM은 전처리과정을 완료한 텍스트들을 분석하여 현재 직무별로 요구되는 역량 중 어떤 역량이 중요한지 알아낸다. CSM이 분석한 결과는 사용자에게 전달되고, 사용자는 CSM의 결과를 언제든 조회할 수 있다. 사용자는 CSM의 결과를 사용하여 취업에 필요한 역량과 관련된 포트폴리오를 작성하거나, 이력서를 온라인에 등록한다. 그림 6은 CSM의 동작 과정을 나타낸다.

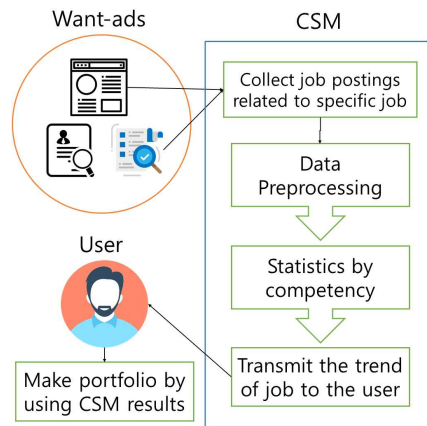


그림 6. CSM의 동작  
Fig. 6. The operation of CSM

### 3.3.1 Data preprocessing

CSM는 텍스트를 분리하고, 분리된 단어들어 어떤 단어인지 판단하는 데이터 전처리과정을 진행한다.

표1. Attribute Table의 예  
Table 1. The example of Attribute Table

No.	Attribute	value
1	job	S/W
2		server
3		Client
4		Web
...		....
...	Capability	Java
...		C
...		C++
...		C#
...		Oracle
...		오라클
...		...
...	Prerequisite	php
...		석사
...		학사
...		박사
...		...
...		졸업 예정
...		신입
n		경력

```

Algorithm1. CSM 직무별 핵심역량 분류
Data Analysis(CT[], AT[], AS[], size_CT, size_AT){
//AT[] is set of the Attributes
//CT[] is set of the Capability Tables
//AS[] is set of the query

for(i=0; i<size_AT; i++){
  for(j=0; j<size_CT; j++){
    AS[i] = add(AS[i], CT[j].sql( "SELECT
      Capability FROM CT[j]
      WHERE CT[j].Attribute = AT[i]
    " ))
  }
}
for(i=0; i<size_AT; i++){
  dtm = TermDocumentMatrix(AS[i])
  //TermDocumentMatrix() is the function that
  calculate the number of occurrences of words in
  each sentence.
  m = as.matrix(dtm)
  v = sort(rowSums(m),decreasing=TRUE)
  d[i] = data.frame(word = names(v), freq=v)
  head(d[i], 10)
}
}
    
```

CSM은 KoNLP 패키지를 사용하여 광고들에 존재하는 텍스트 중 명사를 추출한다. 구인 광고의 텍스트들이 전부 명사로 분리되었다면, CSM은 Attribute Table을 사용하여 각 명사에 대한 속성을 결정한다. Attribute Table이란, 각 명사가 어떤 속성을 지니는지 결정하는 테이블이며, 개발자가 미리 작성해 두어야 하는 테이블이다. 표 1은 Attribute Table의 몇 가지 예를 보여준다.

CSM이 Attribute Table을 이용하여 각 명사에 속성을 부여하면, CSM은 해당 광고에 대한 Capability Table을 생성한다. 그림 7은 데이터 전처리과정의 흐름을 보여준다.

예를 들어 광고 문구 중 'S/W 개발 신입, 2019년 2월 졸업 예정인 학사 또는 석사, Java, Kotlin 을 이용한 서버 개발이 가능하신 분'이라는 문구에 데이터 전처리과정을 적용한다고 가정한다.

우선 KoNLP를 이용하여, 해당 문구의 명사를 추출한다. 나누어진 명사들은 [s/w, 개발, 신입, 2019, 년, 2, 월, 졸업, 예정, 학사, 석사, Java, Kotlin, 이용, 서버, 개발, 가능, 분]이 된다.

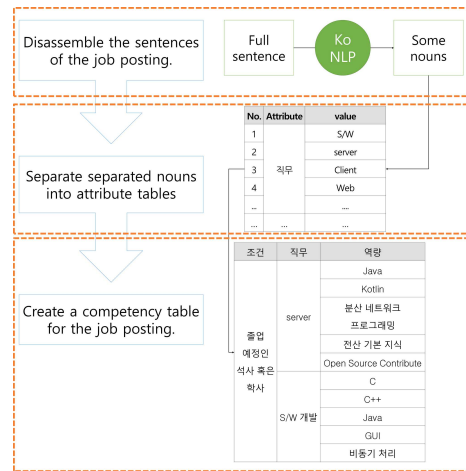


그림 7. 데이터 전처리과정  
Fig. 7. Data preprocessing process

다음으로 각각의 명사를 Attribute Table에 적용한다. Attribute Table에 적용된 명사를 속성별로 구분하면, 직무:[s/w, 개발, 서버], 역량:[Java, Kotlin], 조건:[졸업, 예정, 석사, 학사, 신입, 나머지:[가능, 분,

2019, 년, 2, 월, 이용, 빈으로 나눌 수 있다.

마지막으로, 이 중 나머지를 제외한 직무, 역량, 조건을 사용하여 해당 구인 광고에 대한 Capability Table을 생성한다. 표 2는 생성된 Capability Table이다.

### 3.3.2 Data Analysis

CSM가 충분히 많은 Capability Table을 생성했다면, CSM는 Capability Table들을 사용하여 각 직무에서 핵심적으로 요구되는 역량이 무엇인지 계산한다. 이는 워드 클라우드를 생성하는 과정과 비슷하다. CSM이 직무별 핵심역량을 분류하는 과정은 다음과 같다.

표2. Capability Table의 예  
Table 2. The example of a Capability Table

Prerequisite	job	Capability
졸업 예정인 석사 혹은 학사	server	Java
		Kotlin
		분산 네트워크 프로그래밍
		전산 기본 지식
		Open Source Contribute
	S/W 개발	C
		C++
		Java
		GUI
		비동기 처리

첫째, CSM은 여러 Capability Table에서 직무가 같은 튜플들을 하나의 쿼리로 만든다.

둘째, CSM은 만들어진 쿼리에서 각 역량의 발생 빈도를 계산한다.

셋째, 발생 빈도가 가장 많이 등장한 역량이 해당 직무의 핵심역량이라고 판단하고, 각 역량이 등장한 비중을 원형 혹은 막대 그래프로 시각화한다.

CSM가 직무별 핵심역량을 분류하는 알고리즘은 알고리즘 1과 같다.

알고리즘 1에서, CT[]는 Capability Table의 집합이고, AT[]는 Attribute Table의 집합이다. AS[]는 여러 개의 Capability Table 중 Attribute의 값이 같은

튜플들을 선택하여 생성한 쿼리다. d[]은 각 직무에서 가장 많이 등장한 상위 10개의 역량을 저장한 배열이다. d[]를 이용하여 그래프, 워드 클라우드 등을 생성하여 사용자에게 시각적으로 중요한 정보를 제공할 수 있다. 그림 8은 CSM이 Capability Table로부터 쿼리를 생성하고 쿼리로부터 핵심역량을 계산하여 사용자에게 시각적으로 정보를 제공하는 것을 나타낸다.

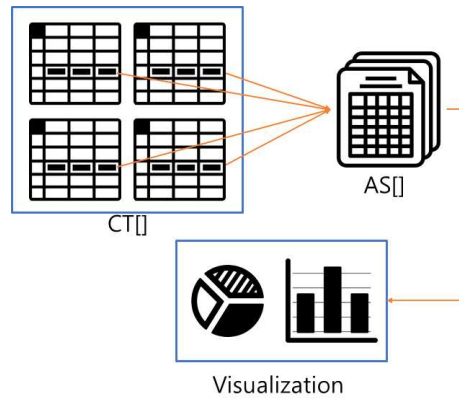


그림 8. Capability Table를 사용한 역량 분류  
Fig. 8. Competency classification using Capability Table

결과적으로 CSM은 웹사이트의 구인 광고들에서 현재 직무별로 중요하게 요구되는 역량이 무엇인지 분석하고, 그것을 시각화하여 사용자에게 전달하여 사용자가 자신의 직무에 맞춰 자신의 역량을 관리하고 개발할 수 있도록 돕는다.

### 3.4. OBBM(Optimize Browser Behavior Module)

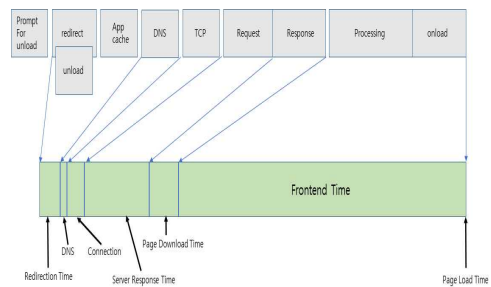


그림 9. 웹 페이지의 동작 과정  
Fig. 9. The process of Web Pages



본 논문은 OOPP의 활용성 및 편리성을 향상하기 위해서 브라우저 처리속도를 향상하는 OBBM (Optimize Browser Behavior Sub-Module)을 제안한다. OBBM은 병렬 검색엔진 모듈인 PSES (Parallel Search Engine Sub-Module), CSS Sprites를 사용하여 이미지 로드를 최적화하는 OILS(Optimized Image Loading Sub-Module)로 이루어진다.

본문에서 제안한 OBBM은 브라우저가 네트워크와 통신을 하는 과정과 통신 후에 페이지를 사용자에게 보여주는 과정의 응답 속도를 개선한다. 최적화 방법을 설명하기 전에 플랫폼의 웹페이지 브라우저가 동작하는 단계를 설명한다.

그림 9는 웹페이지의 동작을 나타낸 다이어그램이다. Prompt for unload는 서비스 이동 단계로 사용자가 웹 페이지 서비스를 이용하다가 다른 주소로 이동할 때 브라우저가 제일 먼저 실행하는 단계로 다른 주소로 이동하기 전에 보고 있던 페이지에서 실행하는데, 모두 브라우저 내부에서 처리하기 때문에 브라우저의 성능과 직결된다. redirect는 리다이렉트 단계로 사용자가 요청한 URL에서 다른 URL로 다시 전송하는 단계다. 리다이렉트를 사용하는 것은 사용자가 의도치 않은 실수로 리다이렉트가 발생했을 때 웹페이지의 성능 향상을 잡아주는 역할을 한다. APP Cache는 애플리케이션 캐시 확인 단계로 플랫폼 브라우저의 캐시에 데이터가 있는지 확인하는 단계다. 전 단계인 리다이렉트 작업을 수행하고 HTTP 요청을 처리하기 위해서 브라우저는 맨 먼저 서버로 요청을 보낸다. 요청에 따른 서버의 응답이 오면 개별 요소가 사용자 PC에 존재하고 있는지 캐시 데이터를 분석한다. 성능 향상을 위해 가장 효과적인 방법은 브라우저와 서버 사이의 통신을 최대한으로 줄이는 것이다. 그러므로 본 플랫폼을 재방문한 사용자에게 좀 더 빠른 응답 속도를 제공하기 위해 애플리케이션 캐시의 활용성을 더욱 강조했다.

DNS ,TCP, Request, Response는 네트워크 통신 단계로 플랫폼의 브라우저가 네트워크와 통신을 하고 웹페이지와 구성요소들을 다운로드한다. 이 네 단계를 거치는 비용이 상당히 크기 때문에, 비용을 줄이면 서비스 성능이 크게 향상된다. 본 논문에서는

Request-Response 사이에서 작동하는 시간을 감소시키기 위해서 PSES(Parallel Search Engine Sub-Module)과 OILS(Optimized Image Loading Sub-Module) 제안하여 Request-Response 시간을 줄인다. 또한, Blocked은 브라우저에서 리소스를 호출하는데 걸리는 시간이다. 이 대기 시간은 브라우저 캐시를 확인하는 시간과 네트워크에 연결하기 전 준비 시간으로 이루어진다.

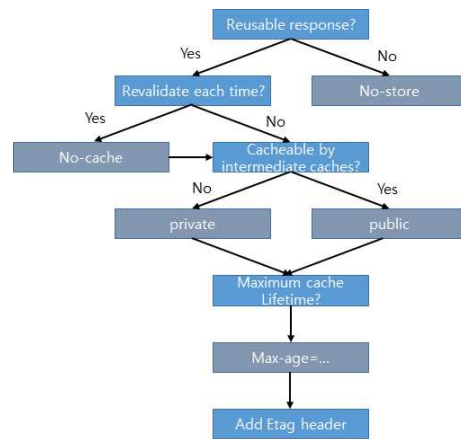


그림 10. 웹 페이지의 캐시 제어 정책  
Fig. 10. Cache Control Policy for Web Pages

동시 연결 개수가 많을수록 대기 시간은 더 짧아지게 된다. DNS(DNS Lookup)는 브라우저에서 IP 주소를 조회할 때 동일한 호스트로 요청하는 요소는 다시 IP 주소를 확인하지 않아 시간이 추가로 걸리지 않는다. 하지만 호스트별 동시 연결 기능은 포기해야 한다. TCP(Connect)는 웹 서버와의 TCP 연결을 생성하는데 필요한 시간이다. 프론트엔드에서는 이 단계에서 걸리는 시간을 줄일 수 없는 백엔드 영역이다. Processing, onload는 브라우저 처리 단계로 다운로드한 웹 페이지와 구성요소들을 이용해서 사용자가 요청한 페이지를 화면에 출력한다.

### 3.4.1 PSES(Parallel Search Engine Sub-Module)

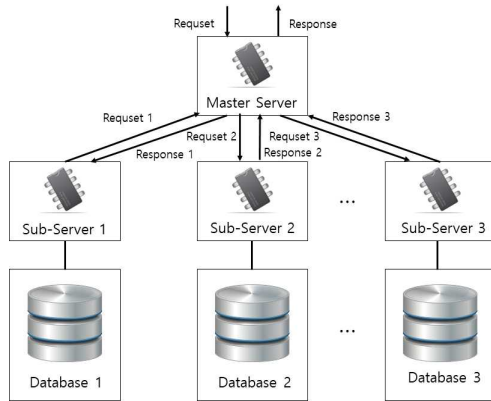


그림 11. 병렬검색 엔진  
Fig. 11. Parallel search engine

검색엔진 최적화를 위해서 PSES(Parallel Search Engine Sub-Module)를 설계한다. 상용화된 검색엔진으로는 Google, Yahoo 등이 있다. 해당 엔진들은 많은 수의 경제적으로 저렴한 PC들을 사용하여 높은 확장성을 지원하는 분산 파일 시스템을 기반으로 구축되었다. 분산 파일 시스템은 데이터를 대규모로 저장하기 적합한 엔진이다. 병렬검색 엔진은 분산 파일 시스템보다 인터페이스를 사용하기 쉽고 결합 발생율이 낮다. 또한, 높은 확장성을 지원하면서 기존에 사용 중인 분산 검색엔진보다 효과적이다. 그림11과 같이 병렬 검색 엔진은 대규모의 데이터에 대한 질의를 병렬 방식으로 처리하는 시스템이다. 병렬 검색엔진은 질의 시 여러 개의 보조 서버(슬레이브 서버)에 분산된 데이터를 병렬방법으로 동시에 검색하고 마스터 서버에서는 보조 서버에서 응답한 데이터를 이용해서 질의에 대한 최종 검색결과를 만들어 낸다. 이때, 보조 서버는 정보 검색 기능을 수행해야 하므로 빠르고 효율적으로 구축해야 한다. 병렬검색 엔진의 아키텍처는 마스터 서버와 보조 서버에서 동작하는 프로그램으로 구성된다. 마스터 서버는 사용자가 입력한 질의를 분할하여 보조 서버에 전달한다. 그리고 보조 서버로부터 전달 받은 질의에 따른 결과를 이용하여 사용자에게 결과를 전달한다. 보조 서버는 마스터 서버로부터 전달받은

질의에 따른 응답 결과를 마스터 서버에 전달한다. 결과적으로, 병렬검색 엔진은 질의 시에도 하나의 질의 및 데이터를 분할하여 작업을 진행함으로써 길어질 수 있는 응답시간을 현저히 줄일 수 있고 대용량 데이터를 검색하고, 검색에 걸리는 시간이 짧다. 또한, 여러 질의를 한꺼번에 처리할 수 있다.

### 3.4.2 OILS(Optimized Image Loading Sub-Module)

CSS Sprite 기법은 여러 이미지를 하나로 만들고 스타일 시트에서 background-position 속성을 원하는 값으로 설정해 필요한 이미지 집합에서 필요한 이미지만 보여주는 기술이다. 그러므로, OILS(Optimized Image Loading Sub-Module)는 여러 개의 이미지를 하나로 합치기 때문에 HTTP 요청 횟수를 줄일 수 있고, 이미지의 컬러 테이블과 같은 메타 데이터를 하나로 합칠 수 있어 파일 크기가 줄어든다. 알고리즘 1은 하나로 합친 CSS 스프라이트 이미지를 사용하기 위한 형식으로 웹 사이트에 사용되는 여러 이미지를 하나의 이미지로 합치고 사용하는 기본적인 알고리즘이다.

#### 알고리즘 2. CSS sprite

```
.menu_list li a
{background:url(http://example.com/some.png)
no-repeat}
.menu_list a.me{background-position:0 0}
.menu_list a.mail{background-position:0 -60px}

.menu_list li a{background:url(http://example.com/some.png) no-repeat}는 CSS sprite를 사용하기 위해서 이미지를 설정한다.
```



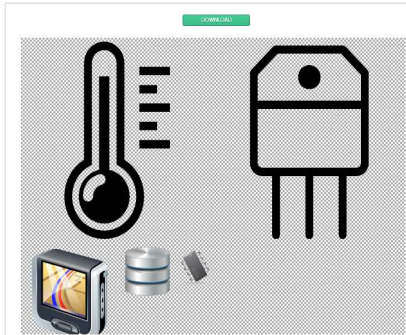


그림 12. CSS Sprite Generator  
Fig. 12. CSS Sprite Generator

그리고 background-position이라는 속성의 왼쪽과 위쪽을 기준으로 보여 줄 이미지를 지정한다. '.menu\_list a.mefbackground -position:0 0'에서는 background-position이라는 이미지 설정에서 좌푯값을 이용하여 왼쪽으로 0 위쪽으로 0 위치에 있는 이미지를 뜻하는 것으로 맨 첫 번째 이미지로 사용한다. '.menu\_list a.mail{back -ground-position:0 -60px}'에서는 첫 번째 이미지를 기준으로 60픽셀 아래에 있는 아이콘이 두 번째 이미지로 사용한다. 그림 12는 OILS를 이용하여 웹 페이지에서 사용하는 그림을 최적화한 예시다.

#### 4. 성능 분석

##### 4.1 CSM 성능검증

CSM의 신뢰성을 측정하기 위하여 본 논문은 두 가지 실험을 실행한다. 두 실험은 3.2GHz 연산 능력과 16GB RAM을 가진 PC에서 실행한다. 첫 번째 실험은 CSM로 생성된 Capability Table과 실제 광고 사이의 유사성을 비교하여 Capability Table의 정확성을 측정한다.

Capability Table은 R언어의 KoNLP를 기준으로 생성되며, 역량을 분류하기 위하여 각종 전문용어는 KoNLP 사전에 등록되어 있다. 광고 페이지는 50개에서 800개로 증가시키면서 정확도를 비교한다.

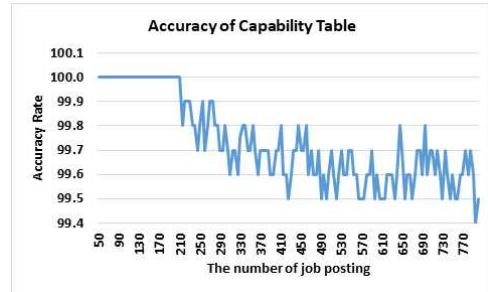


그림 13. Capability Table의 정확도  
Fig. 13. The Accuracy of Capability Table

그림 13은 광고 페이지 수에 따른 Capability Table의 정확도를 보여준다. 광고 페이지 수가 50개부터 약 200개까지는 Capability Table과 실제 광고가 정확하게 일치한다. 광고 페이지 수가 많아질수록 정확도는 점차 감소하지만, 가장 낮은 정확도는 99.4%로 실제 광고와 Capability Table의 차이가 거의 없음을 알 수 있다.

##### 4.2 OBBM 성능검증

본 논문에서 제안한 Optimize Browser Behavior Sub-Module (OBBM)의 성능검증을 위해 최적화 전의 작업시간과 최적화 후의 작업시간을 비교했다. 시뮬레이션은 표 3과 같은 환경에서 진행했다.

표3. OBBM 성능검증의 실험 환경

Table 3. Experimental environment for the OBBM performanceverification

OS	Window 10 Education
CPU	Intel Core i5 4590
GPU	NVIDIA GeForce GTX 1050
Memory	DDR3 8GB-12800
Browser	Internet Explorer 11

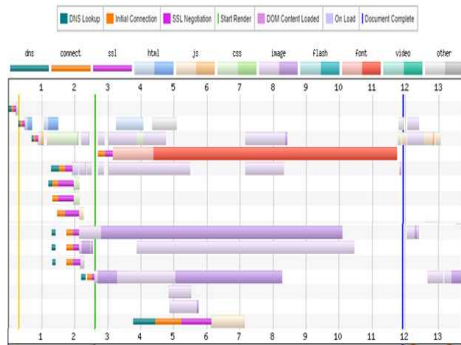


그림 14. 최적화 전 작업시간  
Fig. 14. Working time before optimization

그림 14는 OBBM을 이용하여 최적화 전의 작업시간을 나타낸다. 플랫폼의 페이지 로딩에 걸리는 시간이 11.918s가 소모되었다. js, dns, ssl은 평균적으로 0.8s의 시간을 소모하였다. 그리고 가장 많은 시간은 font를 로딩하는데 걸리는 시간으로 약 9s의 시간이 소모되었다. 또한, font 다음으로 작업시간이 오래 걸린 image 같은 경우는 약 8s 정도 소모되었다. 하지만 font와는 달리 한 번의 작업을 수행하지 않고 여러 image를 로딩하므로 많은 다중작업이 필요로 한다. 작업에 소모되는 시간 대부분이 font, image 파일을 불러오는데 걸리는 시간이다. 그러므로 두 가지 요소의 작업시간을 줄이는 최적화 작업을 수행하면 작업시간을 크게 줄일 수 있다.

그림 15는 OBBM을 이용하여 최적화 작업 후의 플랫폼 작업시간을 나타낸다. 최적화 작업 전에는 font, image 작업에서 가장 많은 시간이 소모되었다.

하지만 최적화 후에는 font의 작업 수행시간은 거의 영향을 미치지 않을 만큼 감소하였다. 또한, image를 로딩하는데 걸리는 시간이 약 3s 정도로 감소했다.

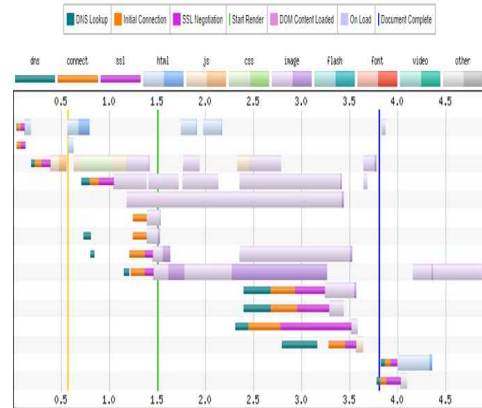


그림 15. 최적화 후 작업시간  
Fig. 15. Working time after optimization

최적화 전의 image로딩 시간인 8s와 비교했을 때 62.5%의 시간이 감소했다. image를 다중처리하는 작업량은 증가했으나, 전체적인 작업시간은 크게 줄어들었다. 또한, 플랫폼 전체의 작업에 필요한 소모시간을 비교했을 때 최적화 전에는 11.918s의 시간이 필요했으나, 최적화 작업 후에는 전체 작업이 3.785s에 완료되었다. 최적화 전과 후를 비교했을 때 전체 작업시간은 약 68.37%가 감소 되었다.

## 5. 결론

본 논문은 구직자가 자신에게 맞는 직무의 핵심 역할이 무엇인지 편리하게 확인하고, 자신만의 포트폴리오를 작성 및 관리할 수 있는 OOPP를 제안했다. OOPP의 특징은 다음과 같다.

첫째, 직업정보 사이트의 구인 광고 정보를 스프레드시트에 저장한다.

둘째, 저장된 정보를 정규화하고 직무별 주요역량이 무엇인지 분석한다.

셋째, 웹페이지 최적화를 통하여 분석한 정보를 신속하게 사용자에게 제공한다.

넷째, CSM을 이용하여 분석한 구인 광고는 실제 광고와 최대 100%, 최소 99.4% 일치한다.

다섯째, OBBM을 웹페이지에 적용했을 때 웹페이지의 처리속도는 OBBM을 적용하기 전보다 68.37% 향상되었다.

여섯째, OOPP는 현재 직업정보 사이트의 구인 광고를 정확하게 분석하여 사용자가 분석한 결과를 웹페이지에서 신속하게 조회할 수 있다.

## REFERENCES

- [1] Microsoft, "Linkedin", <https://kr.linkedin.com/>, 2018
- [2] Mina Choi, Hyelan Roh, Myungsook Kim, "The Development of e-Teaching and Learning Portfolio Framework in University Education.", *The Journal of Educational Information and Media*, Vol.11, No.2, pp.147-165, 2005.
- [3] Sunghyun Han, "An Implementation of Web-based E-portfolio Management System.", *Journal of The Korea Society of Computer and Information*, Vol.13, No.5, pp.37-44, 2008.
- [4] Hyugil Kwon, "Technological Approach to Performance Assessment: Development of a Web-Based Electronic Portfolio", *Journal of Educational Technology*, Vol.18, No.1, pp.51-78, 2002.
- [5] Changhui Choi, "Optimizing an Insurance Company's Portfolio with a Consideration of the RBC constraint.", *Journal of insurance and finance*, Vol.25, No.2, pp.3-32, 2014.
- [6] Dong-Jin Park, Yun-Jae Lee, Jin-Gyoung Kim, "A Case Study of e-Portfolio Framework Based K-portfolio Implementation for Improving College Students' Competency.", *The Journal of Korean Association of Computer Education*, Vol.17, No.2, pp.87-95, 2014.
- [7] Woonju Cho, Ilsun Choi, Eunyoung Kim, "Developing Rubrics for Assessment for a Teaching Portfolio in Kindergarten", *Early Childhood Education Research & Review*, Vol.15, No.6, pp.213-235, 2011.
- [8] JeWoo Jason Sun, Jeongyoen Yoon, Seongjoo Song, "Copula-normal inverse Gaussian model for portfolio risk assessment.", *The Journal of Risk Management*, Vol.29, No.1, pp.22-36, 2018.
- [9] Dohoon Kim, "A Strategy for Optimal Production Management of Multi-Species Fisheries using a Portfolio Approach.", *The Journal of Fisheries Business Administration*, Vol.45, No.1, pp.109-119, 2014.
- [10] M.H. Everts, H. Bekker, J.B.T.M. Roerdink, T. Isenberg, "Depth-Dependent Halos: Illustrative Rendering of Dense Line Data.", *IEEE Transactions on Visualization and Computer Graphics*, Vol.15, No.6 pp.1299-1306, 2009.
- [11] Seung Taek Ryoo, "Cartoon Rendering For Cartoonly Surface Detail", *Journal of the Korean society for computer game*, Vol.11, No.11, pp.13-19, 2007.
- [12] Taeho Kim, Jinah Pack, "Effective Volume Rendering and Virtual Staining Framework for Visualizing 3D Cell Image Data.", *Journal of The Korea Computer Graphics Society*, Vol.24, No.1, pp.9-17, 2018.
- [13] Heekyung Yang, Kyungha Min, "A Realistic Pencil Rendering Technique for Game Characters.", *Journal of the Korean society for computer game*, Vol.23, No.23, pp.49-59, 2010
- [14] MiRa Park, KeyHo Park, JaeSeong Ahn, "Design and Implementation of a Computing Environment for Geovisual Analytics Using HTML5.", *Journal of the Korean Association of Geographic Information Studies*, Vol.14, No.4, pp.44-53, 2011.
- [15] Andrea Forte, Thomas H. Park, Brian Dorn, "An Analysis of HTML and CSS Syntax Errors in a Web Development Course.", Vol.15, No.1, pp.1-21, 2015.
- [16] Zhongcheng Lei, Hong Zhou, Wenshan Hu, Qijun Deng, Dongguo Zhou, Zhi-Wei Liu, "HTML5-Based 3-D Online Control Laboratory With Virtual Interactive Wiring Practice.", *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, Vol14, No.6, pp.2473-2483, 2018.
- [17] Heewon Jeon, "Introduction to KoNLP API", <https://cran.r-project.org/web/packages/KoNLP/vignettes/KoNLP-API.html>, 2016
- [18] JongHwa Lee, "Distributed Processing Systems Design For Web Mining Optimization Through Realtime Crawling", 2017
- [19] Woojin Park, Kiyun Yu, "Spatial Clustering Analysis based on Text Mining of Location-Based Social Media Data", *The Korea Society For Geospatial Information Science*, Vol. 23, No. 2, pp.89-96, 2015
- [20] Gilseung Ahn, Minji Seo, Sun Hur, "Development of Accident Classification Model and Ontology for Effective Industrial Accident Analysis based on Textmining", *Journal of The Korean Society of Safety*, Vol. 32, No. 5, pp.179-185, 2017

---

저자약력

---

정보근(Bogeun Jung)

[정회원]



- 2004년 2월 : 가톨릭관동대학교 컴퓨터공학과(공학학사)
- 2006년 8월 : 가톨릭관동대학교 컴퓨터교육과(컴퓨터교육석사)
- 2018년 10월 현재 : 가톨릭관동대학교 컴퓨터공학과 박사과정

<관심분야>

빅데이터, SNS, Web, 사물인터넷

박진욱(Jimuk Park)

[정회원]



- 2013년 2월 : 가톨릭관동대학교 체육교육과(체육학사)
- 2015년 8월 : 가톨릭관동대학교 의료경영학과(의료경영학석사)
- 2018년 10월 현재 : 가톨릭관동대학교 컴퓨터공학과 박사과정

<관심분야>

헬스케어, 빅데이터, 네트워크보안, 사물인터넷

이 병 관(ByungKwan Lee)

[정회원]



- 1979년 2월 : 부산대학교 기계설계학과(공학사)
- 1986년 2월 : 중앙대학교 전자계산학과(이학석사)
- 1990년 2월 : 중앙대학교 전자계산학과(공학박사)
- 1988년 3월 ~ 현재 : 가톨릭관동대학교 소프트웨어학과 교수

<관심분야>

네트워크보안, 빅데이터, 데이터마이닝, 사물인터넷