**<별첨8>**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4차산업혁명을 선도하는 CODE형 SW 인재 양성  Capstone Design(종합설계) 결과보고서 | | | | | | | | | |
| **교과목명** | | **스마트IoT캡스톤디자인** | | | **학부(과)명** | | **스마트IoT** | | |
| **작 품 명** | | 웹크롤링을 이용한 IT취업 플랫폼 개발(Job Hound) | | | | | | | |
| **팀 명** | | 캡스톤코인 | | | | | | | |
| **지도교수** | | 박지현 교수님 | | | | | | | |
| **참여**  **학생** | **대표** | 소속학과 | | 학번 | | 성명 | | 연락처 | |
| 스마트IoT | | 20145153 | | 이종학 | | 010-4052-1024 | |
| **팀원** | 소속학과 | | 학번 | | 성명 | | 연락처 | |
| 스마트IoT | | 20145145 | | 유홍재 | | 010-2740-8430 | |
| 스마트IoT | | 20145148 | | 이민준 | | 010-9427-9398 | |
| 스마트IoT | | 20167207 | | ZHENG JUN | | 010-4327-0415 | |
|  | |  | |  | |  | |
| **참여 분야** | | □ SW융합대학 □ SW융합/연계 전공 | | | | | | | |
| **참여기업** | | 없음 | | | | | | | |
| **지식재산권** | | **유형** | □ 특허 □ 실용신안 □ 상표 □ 디자인 □ 기타( ) | | | | | | |
| **명칭(출원명)** | | | | | | | **번호** |
|  | | | | | | |  |
| **과제 목적** | | | | | | | | | |
| **프로젝트 배경 및 목적**  EMB000035ec546c블록체인, IoT, 알파고로부터 비롯된 AI에 관한 관심도가 늘어나면서 전세계적으로 IT인력의 수요는 급증하고, IT전공자 뿐만 아니라 IT비전공자들도 IT업계에 취직하기 위해서 복수전공과 전과 등을 하며 노력하고 있다. 하지만 취업희망자들에게는 너무나도 많은 IT기업의 채용 공고로 신뢰도가 부족한 상황이다. 삼성 및 구글 같은 대기업은 정보가 많고 뉴스와 광고를 비롯한 미디어매체에도 많이 노출되는 만큼 정보를 쉽게 얻을 수 있다. 중견기업은 그나마 사정이 나을지도 모르지만 중소기업의 경우에는 홈페이지가 없는 경우도 있고 회사 정보를 얻기 쉽지 않다. 그리고 이 기업이 정확하게 무슨 일을 하는지 알 수 없는 경우도 많고 취업준비생들이 따지는 연봉, 복지 등 회사선택에 중요한 요소들도 알기가 어렵다. 취업시장에서는 정보가 곧 경쟁력이다. 그로 인해 수많은 취업정보사이트 들이 현재 개설되어 있으며, 다양한 기업들의 채용 정보, 합격 수기 등이 게시되어 있다. 그러나 이러한 사이트들을 이용함에 있어서 대부분의 사람들이 불편함을 느낀다.  **그림1. 가짜 구인정보 관련 기사**  <http://imnews.imbc.com/replay/2018/nwtoday/article/4772685_22669.html>  1. 첫째로, 가짜 구인정보 및 광고 글 때문이다. 취업정보 사이트에는 대기업, 유망 기업 뿐만이 아니라 소규모 스타트업 기업들을 포함해서 채용에 대한 정보가 올라온다. 그 중, 취업정보 사이트에 광고비를 내고 광고하는 기업들도 있을 뿐 더러, 허위 구인광고를 내세워 가짜 정보를 제공함으로써 사용자들의 개인정보를 빼내고 시간낭비를 유도한다.  EMB000035ec546d  2. 둘째로, 취업정보 사이트가 복잡하여 원하는 정보를 찾는데 시간이 많이 소모되는 것이다. 대부분의 취업정보 사이트에 너무나도 많은 기업들에 대한 정보가 올라와 있고, 정작 사용자가 원하는 정보를 찾으려면 사이트 사용에 익숙해져야 한다. 또한 원하는 정보가 올라왔음에도 불구하고 알림을 제대로 받지 못하거나, 돈을 결제해야 받을 수 있는 경우도  **그림2. 취업정보사이트의 불편한 점 통계(출처 : 대학내일)**  EMB000035ec546f생긴다. 이로 인해 구인정보 데이터를 가공하고,  사용자로 하여금 명확하게 바로 받아볼 수 있도록 해야 한다.  **그림3. 취준생들의 해외 취업에 대한 관심**  3. 마지막으로, 현재 대부분의 취업정보 사이트들은 해외취업에 대한 정보 제공이 막연하고 부족하다. 최근 그림 3와 같이 국내 취업이 힘들어 짐에 따라 해외취업에 관심이 있는 취업준비생들이 매우 많아지고 있다. 그러나, 해외취업에 성공하는 사례는 많지 않다. 조사 결과, 해외취업준비생 중 실제 해외 취업에 성공하는 사례는 22% 밖에 되지 않았다. 또한, 해외취업이 됐다 하더라도 그림5 과 같이 원했던 업무와 다른 일을 하게 되는 경우가 많다. 이들이 해외취업 준 비를 어려워하는 이유 중 ‘해외취업 정보를 얻기 힘들다’ 가 58.8%로 가장 높았으며, ‘어학능력 향상이 어렵다’ 가 43.5%로 그 뒤를 따랐다.이러한 결과들을 보았을 때, 현재 해외취업에 대한 가장 큰 문제점은 정보 제공의 부족이다. 정보를 얻기 위한 또 다른 방법은 직접 해외채용박람회에 참석하는 것이다. 하지만 서울 및 부산 등 대도시에서 주로 실시되기 때문에 지방에 사는 해외취업희망자들은 접근성이 떨어진다. 고로 본 프로젝트에서 기존 취업정보사이트들의 막연하고 진실되지 않은 해외취업 정보 제공을 해결하고자 하고 국내외 IT기업의 필요한 정보들만 추출해서 정확하고 신속하게 알림으로써, 사용자로 하여금 취업 시장 정보 탐색의 불편함을 줄여주고, 취업 성공의 목표를 달성하게 하는 것이다.  EMB000035ec5470  **그림 4. 해외 취업 성공률**  EMB000035ec5472  **그림5 청년 해외취업의 실태**  EMB000035ec5471  **그림6. 해외취업 준비의 어려운 점 설문결과** | | | | | | | | | |
| **과제 내용** | | | | | | | | | |
| **프로그램 제작 방향**  위와 같이 기존 채용정보 플랫폼의 문제를 해곃하고, 한단계 더 나아가고자 본 프로젝트는 IT 기업 취업 플랫폼 “잡하운드”를 제시한다. 이는 IT 취업을 희망하는 사용자들에게 신뢰성과 편리성을 제공하고자 한다. 초기에 IT 취업준비생에게 국내외 IT 기업 중 신뢰되고, 유망한 기업 800개 내외를 선정하였다. 그러나 제작 기간상 현재 사용자들에게 제공하는 기업들의 정보는 선정한 목록 중 8개이며, 본 아이디어의 핵심을 구현하는데 큰 지장은 없다. 다음은 선정한 기업들의 목록에 대한 설명이다.  **1) 국내 상위 100대 IT 기업 : IT서비스기업 매출 순위 상위 100개**  **그림. IT서비스기업 매출 순위표**  위 자료는 한국 IT 서비스 산업 협회에서 제공하는 국내 IT 기업 매출 순위를 순서대로 나열한 자료이다. 이중 상위 100대 기업들은 높은 자본력을 이용해 우수한 연봉을 제공하고, 더 나은 환경에서 일할 수 있도록 좋은 복지 또한 제공해준다. 본 프로젝트는 상위 100대 기업의 정보가 IT취업에 관심을 가지고 있는 사람들에게 있어 가치 있는 정보라 판단하여 선정했다.  **2) 국내 유망 400대 IT 기업 : 한국벤처캐피탈협회 인증 기업**    **그림. 한국벤처캐피탈협회 인증 기업 목록**  국내 상위 100대기업 외에 한국벤처캐피탈협회에서 지정한 국내 유망한 벤처기업 400대를 선정했다. 한국 벤처캐피탈 협회에서 지정한 유망한 기업의 기준은 투자 받은 자본금이 10%이상, 투자 금액 5천만원 이상인 중소기업들이다. 또한 이 기업들은 직원 개개인에 대한 철저한 관리, 우수한 복지정책과 뛰어난 단합력으로 근무 환경이 좋고 이적시에도 풍부한 경력이 된다. 이러한 IT 기업들의 정보를 취업준비생들에게 제공하여 기존에 알지 못했던 우수한 기업에 지원할 수 있도록 하고, 국내 중소기업 생태계를 활성화 시키고자 한다.  **3) 중국 100대 IT 기업: 하이테크기업인증 받은 100대 IT 기업**  **그림. 중국 하이테크기업인증 취득 기업 목록**  본 프로젝트는 해외 IT취업을 희망하는 취업준비생들을 위해 해외에 있는 IT기업을 선정했다. 그 중, 중국의 하이테크기업인증 기업들은 정부의 기준에 따라서 IT 분야에 우수한 성과를 낸 기업들이다. 하이테크인증 기업은 기술적 인원이 부족하여 모집을 많이 하고, 기업에서 제공하는 다양한 혜택이 국내 대기업에 버금간다. 국내에 잘 알려지지 않은 중국 IT기업에 대한 정보를 하이테크기업인증을 통해 전달하고자 한다.  **4) 미국 100대 IT 기업: Forbes 선정 Top 100 Digital Companies**  **그림. 미국 Forbes 선정 Top 100 Digital Companies 목록**  미국 취업의 경우, Forbes에서 매년 주식 실적, 매출, 이익 및 자산성장을 기준으로 순위를 매겨 제공하는 상위 100대 IT 기업을 선정했다. 미국 진출을 함에 있어 보다 우수하고, 세계적인 기업에 취업을 장려하고, 만족스러운 미국 취업 보장하기 위해 위와 같이 선정하여 해외 취업률을 좀 더 높이고자 한다.  **5) 일본 100대 IT 기업: 졸업자 취업 인기 랭킹 100대 기업**  **그림 4. 일본 졸업자 취업 인기 기업 랭킹**  일본의 경우, 매년 졸업예정 학생을 대상으로 조사한 인기 기업 랭킹 중 11가지의 유형별로 8~10위까지 선정하였다. 이는 일본 현지 학생들의 선호도로 조사된 만큼 믿음 직한 도표로서, 일본 취업준비생들에게 신뢰 되는 정보를 제공한다.  따라서 위와 같은 IT 기업들을 선정하였고, 이 중 실제로 프로젝트 구현에 활용한 기업은 국내 대기업인 삼성, 벤처기업인 원더풀 플랫폼이 있다. 그리고 중국 기업인 샤오미, didi, vivo, 미국 기업인 Google, Facebook, Cisco 가 있다. 이 기업들의 정보를 프로젝트의 핵심 기술인 크롤링을 이용하여 사용자들에게 실시간으로 전달한다.  **구성 및 구현 설명**  본 프로젝트에서 선정한 기업들을 어떻게, 사용자들에게 제공할 것인지 구상을 한 결과 아래와 같은 시스템 구성도를 구상하였다. 시스템 구성도의 목적은 웹크롤링을 이용해 채용 정보를 사용자들에게 실시간으로 전달하는 것이다. 핵심 요소인 크롤링 서버, 데이터베이스, 웹서버, 클라이언트에 대해 순서대로 설명하겠다.    **프로젝트 시스템 구성도**  **크롤링**  EMB00004050649b첫번째 핵심요소는 웹크롤러이다. 웹크롤러란 링크를 따라다니며 페이지에서 원하는 정보를 찾아 기록하고 색인하는 것이다. 본 프로젝트는 크롤러를 이용해 기업들에 대한 정보를 탐색한다.  크롤러의 핵심 기능은 Selenium과 Beautifusoup이 있다. BeautifulSoup은 웹의 정보를 읽고 추출하는 역할, 그리고 Selenium은 크롤링일 할 수 있는 기능을 제공하는 역할이다. 이를 이용해 웹의 HTML 기록된 정보를 가져온다.    🡨현재 페이지의 소스코드를 전부 긁어옴  🡨데이터베이스 연결  🡨필요한 내용 크롤링  🡨데이터베이스에 저장  그림. 파이썬 크롤링 코드  EMB00004050649c  **그림. HTML 코드 분석**    **그림. 크롤링의 플로우 차트**  다음은 크롤링의 플로우차트 이다. 이는 각 기업의 채용정보에 대한 크롤링 함수를 구현하고, 선정기업들 목록을 일정 주기로 크롤링 통해 업데이트한다. 그리고 크롤링한 데이터를 데이터베이스에 저장한다.  **데이터베이스**  **그림. 서버에 저장된 데이터베이스**  다음은 데이터베이스이다. 데이터베이스는 크게 두 분류로 나눠진다. 각 기업들의 테이블과, 총 기업들의 목록에 대한 테이블 두가지가 존재한다. 각 종류의 테이블에 해당하는 정보를 저장한다. 데이터베이스 플로우 차트를 참고하면, 크롤링한 데이터와 데이터베이스의 내용을 먼저 비교하고, 이것이 최신 내용 일 때 만 업데이트하여 사용자에게 전달한다.  **그림. 데이터베이스 플로우 차트**  .    **그림. 삼성기업 크롤링 데이터베이스**  **서버**  본 시스템의 세 번째 핵심은 서버이다. 서버가 필요한 이유는 클라이언트가 DB에 직접 접근하면 정보가 유출되는 치명적인 문제가 발생할 수 있다. 따라서 프로그램 관리 및 유지보수, Client와 DB의 상호작용을 중계해 줄 서버가 필요하여 구축하였다.  **그림. 서버의 역할**    **그림. 서버의 PHP파일**  데이터베이스에서 서버, 서버에서 클라이언트로 또는 클라이언트에서 서버, 서버에서 데이터베이스로 데이터의 이동을 중계해주는 php파일이다. php파일의 구성은 기업데이터를 처리해주는 company.php파일, 사용자 개인정보를 처리하는 login.php 마지막으로, 국내 대기업, 벤처기업, 해외기업 등의 데이터를 처리해주는 cisco.php, didi.php등이다.  Table [Company]  company.php  Table[기업명]  facebook.php, google.php ...  그 외 실제로 구현하지는 못했으나 다음과 같은 기능을 서버에 추가하려 하였다.   1. 서버에서 클라이언트로 알림서비스 제공. 2. 키워드 검색시 데이터베이스의 튜플과 문자열 비교를 하기 위해 서버의 중계역할이 필요. 3. 사용자가 즐겨찾기한 기업목록 데이터베이스에 접근하기 위해 서버의 중계역할이 필요 .   **클라이언트**  **초기 클라이언트 구성도**  마지막으로 클라이언트이다. 클라이언트는 안드로이드 스튜디오를 이용해 제작하였고, 웹서버와 HTTP 통신하며 사용자가 원하는 정보를 가져온다. 왼쪽 자료는 초기 클라이언트 구성도로 사용자가 원하는 데이터를 가공해서 보여주는 것과, 실시간으로 크롤링한 정보를 보기 쉽게 나타내는것이 목적이었다.  **기업명 / 분야**  **직원수 /평균초봉**  **- 기업의 채용 공고에서 각 공고당 최대 5가지 정보를 전달**  **EX)제목, 경력구분, 세부분야, 날짜 등**  **클라이언트 CISCO 채용 공고목록**  **클릭시 링크 이동**  **클라이언트 미국 기업목록**  위 그림은 구 체적인 클라이언트 모습이다. 처음에 각 기업의 기업명, 분야, 직원수, 평균 초봉을 보여주고 기업을 누르면 해당 기업의 채용 공고가 나타난다. 공고는 가공한 데이터를 총 5줄로 표현하며, 기업마다 다르지만 대체적으로 제목, 경력구분, 세부분야, 날짜, 등을 보여준다. 또한 공고를 클릭하게 되면 해당 공고의 링크로 이동하게 된다.  **클라이언트 로테이션 기업**  - 매일 국내, 해외별 로테이션 기업 제공    - 다양한 기업들 접할 기회 제공  - 로테이션 기업 지원시 경품 추첨 등  혜택 제공  **클라이언트 국내기업 로테이션**  위 클라이언트 기능 이외에도 ‘금일의 로테이션 기업’ 이라는 기능을 추가하였다. 이는평소 접하지 못한 다양한 기업들을 사용자들에게 알리고자 매일마다 랜덤으로 기업들을 선정하여 사용자들에게 제공한다. 무작위 선정은 랜덤의 시드값을 하루 단위로 주어 국내, 해외별 5개를 선정하도록 했고, 이 기업들에 지원시 경품 추첨 등과 같은 해턕을 줄 예정이다.  이와 같은 구성을 통해 본 프로젝트의 어플리케이션 ‘잡하운드’ 제작을 완료했다. | | | | | | | | | |
| **활용 방안 및 기대효과** | | | | | | | | | |
| **프로젝트의 기대효과 및 활용방안**  본 프로젝트를 통해서 기대할 수 있는 효과는 다음과 같다.  1. 사용자는 국내외 유망 회사의 성격을 명확하게 파악할 수 있으므로 사용자의 취향에 맞는 기업을 선택하는 것이 간단해지고, 목표의식을 뚜렷하게 부여해준다.  2. 사용자들이 취업을 원하는 기업의 정보를 실시간으로 전달받아 취업 경쟁에 있어 우위를 점할 수 있다.  3. 허위 및 가짜 정보, 광고 등이 없으므로 사용자들로 하여금 정보 탐색의 불편함을 없애주고, 시간낭비를 없애준다.  본 프로젝트는 까다로운 기준으로 선발한 신뢰되는 기업들에 대한 정보만을 제공한다. 그리고 취업준비를 하는 대학생들과 졸업한 취업준비생들이 어플을 보고 해당 IT 기업의 취업정보를 손쉽게 볼 수 있다. | | | | | | | | | |