

JobBoost

1. Conceptualization Document



22112022

이광은

nkwang0311@naver.com

[Revision history]

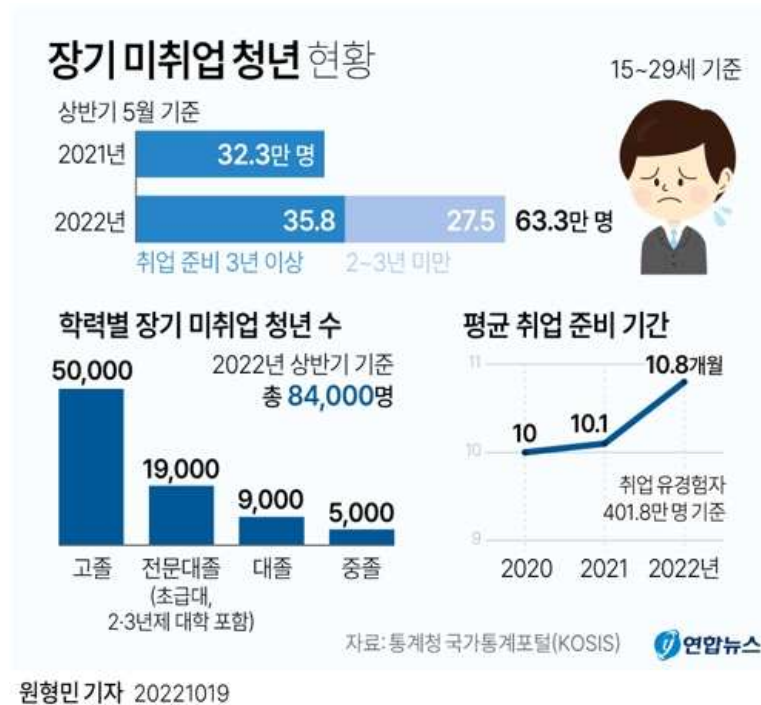
| Revision date | Version # | Description | Author |
|---------------|-----------|---------------------|--------|
| 03/29/2023 | 0.01 | First Documentation | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

= Contents =

| | |
|---------------------------------|----|
| 1. Business purpose | 3 |
| 2. System context diagram | 5 |
| 3. Use case list | 6 |
| 4. Concept of operation | 8 |
| 5. Problem statement | 12 |
| 6. Glossary | 14 |
| 7. References | 15 |

1. Business purpose

1.1 Project background & motivation



[그림 1] 취업 시장 통계자료

위의 [그림 1]¹⁾을 통해 현재 취업시장 상황을 알 수 있다. 작년에는 재작년에 비해 취준생의 수가 증가하였고, 구직난이 심해지고 있다. 앞으로도 계속해서 취업 난이도가 상승될 것으로 보여진다. 또한, 취업준비를 위해 평균 취업 준비 기간이 10개월 정도로 긴 시간을 쏟고 있다. 취준생들이 취업 정보를 얻을 수 있는 플랫폼은 대표적으로 잡플래닛과 잡코리아가 있다. 그 중 잡플래닛은 대학 제휴를 통하여 리뷰를 작성하지 않고, 학교 계정을 통하여 정보를 얻을 수 있게 되어있다. 잡플래닛 웹페이지를 살펴보면 다른 사용자들이 작성한 리뷰를 각각 볼 수 있게 되어있다. 이러한 방법으로는 회사의 장점을 모아보고, 단점만을 모아보는 등 원하는 정보만을 정리하여 보기 불편하다. 면접 내용도 마찬가지이다. 면접에 관한 글을 일일이 저장하여 면접을 준비하기는 힘들 것이다. 물론, 멤버십을 이용하면 기업 간 요약정보와 한눈에 비교하기 편한 ui를 제공한다. 그럼에도 회사별로 두세 개의 대표적인 장단점만 볼 수 있다는 한계가 있다. 더군다나 취준생은 하나의 회사만을 준비하는 것이 아니다. 잡코리아 2021년 통계²⁾에 따르면 2021년 신입 구직자들은 올 하반기에 평균 6.1회 입사 지원해 1.4회 서류 합격한 것으로 집계됐다고 한다. 즉, 취업 준비 중에 번거로운 작업을 하여 시간을 낭비할 만큼 시간이 여유롭지 않다. 따라서 사용자가 원하는 기업 정보만을 볼 수 있도록 하고, 크롤링

을 통한 면접 기출집 제작 기능을 지원함으로써 쉽게 취업 준비를 할 수 있도록 편의성을 제공하고, 취준생의 시간을 아껴주고자 이러한 프로그램을 기획하게 되었다.

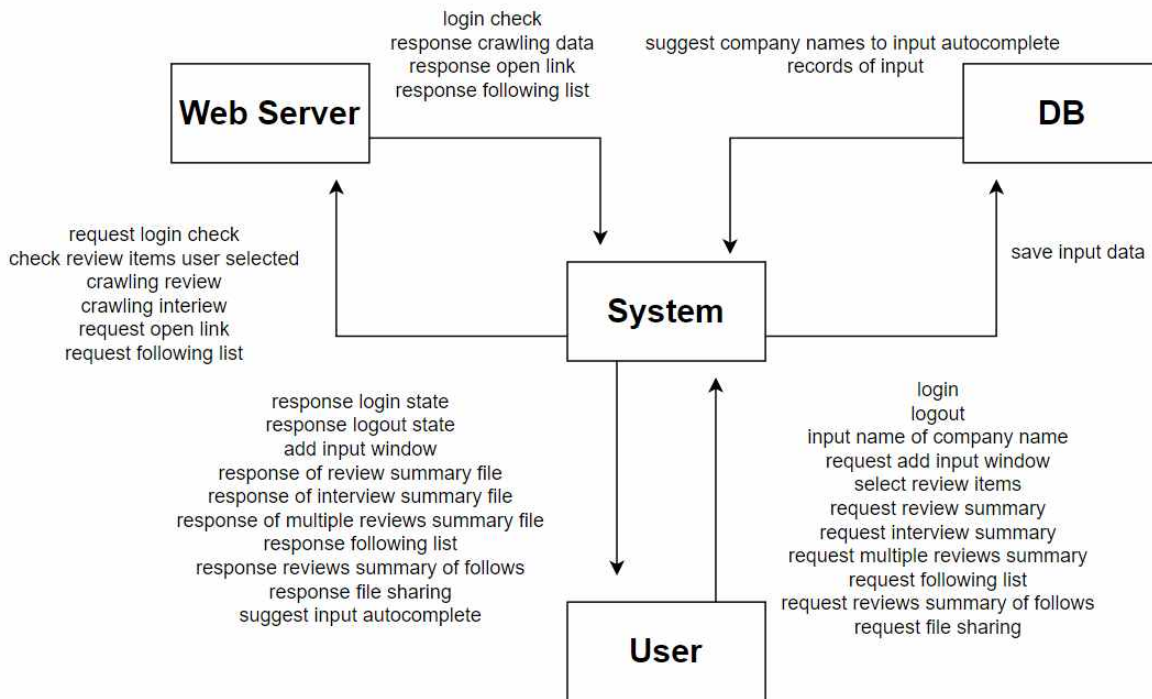
1.2 Goals

- 파이썬 웹 크롤링을 통한 서비스 개발
- 잡플래닛 리뷰를 한눈에 보는 것에 불편함을 해소
- 취업 준비 효율성 증가

1.3 TargetMarget

잡플래닛을 사용하는 사용자 혹은 취준생

2. System context diagram



- login : 로그인
- request login check : 로그인 ID/PW 체크
- logout: 로그아웃
- add input window : 입력창 추가 (여러 회사명 입력)
- select review items : 리뷰 요약에 위한 세부 항목 선택
- response of review summary file : 리뷰 요약 파일 출력
- response of interview summary file : 면접 정보 요약 파일 출력
- response of multiple reviews summary file : 여러 리뷰 요약 파일 출력
- following list : 팔로잉 목록
- reviews summary of follows : 팔로잉 회사들의 리뷰 요약
- file sharing : 파일 공유
- suggest input autocomplete : 입력어 자동 완성
- suggest company names to input autocomplete : 입력어 자동 완성을 위한 회사명 제시
- records of input : 입력 기록 저장
- crawling review : 리뷰 정보 크롤링
- crawling interview : 면접 정보 크롤링
- open link : 하이퍼링크 실행

3. Use case list

3.1 로그인

| | |
|-------------|--|
| Actor | User, WebServer |
| Description | 사용자가 자신의 아이디로 로그인한다. 시스템을 잡플래닛 웹서버에 로그인 정보를 주고 로그인 성공 여부를 받아 반환한다. |

3.2 로그아웃

| | |
|-------------|----------------------|
| Actor | User, WebServer |
| Description | 로그아웃하고 초기 화면으로 돌아간다. |

3.3 회사명 입력

| | |
|-------------|------------------------------|
| Actor | User |
| Description | 유저가 자신이 찾고 싶은 기업의 회사명을 입력한다. |

3.4 원하는 정보 항목 선택

| | |
|-------------|--|
| Actor | User |
| Description | 리뷰의 세부 항목인 작성날짜, 별점, 장점, 단점, 경영진에게 바라는 점 등 크롤링 할 항목을 선택한다. |

3.5 리뷰 요약

| | |
|-------------|---|
| Actor | User, WebServer |
| Description | 입력한 하나의 회사에 대해 선택한 정보들(장점 단점 등)을 모아 하나의 엑셀 파일로 요약하여 저장한다. |

3.6 면접 기출 정리

| | |
|-------------|-------------------------------|
| Actor | User, WebServer |
| Description | 면접 질문을 문서 파일로 정리해서 추출하여 저장한다. |

3.7 파일 공유

| | |
|-------------|-----------------------------|
| Actor | User |
| Description | 요약된 엑셀 파일이나 문서 파일을 바로 공유한다. |

3.8 여러 회사명 입력

| | |
|-------------|--|
| Actor | User |
| Description | 한 번에 여러 회사를 검색하고 싶은 경우, 회사명 검색 창을 추가하여 여러 회사명을 입력하도록 한다. |

3.9 여러 회사의 리뷰 요약

| | |
|-------------|---|
| Actor | User, WebServer |
| Description | 여러 회사명을 입력한 경우, 회사별로 리뷰 요약, 면접정보 통계 (난이도, 결과) 등 정보를 하나의 엑셀 파일로 요약하여 저장한다. |

3.10 팔로잉 목록 확인

| | |
|-------------|---|
| Actor | User, WebServer |
| Description | 로그인한 사용자의 잡플래닛 계정 내에 정보 중 팔로잉하는 회사 목록 불러온다. |

3.11 팔로잉 중인 회사들의 리뷰 요약

| | |
|-------------|---|
| Actor | User, WebServer |
| Description | 팔로잉 목록을 여러 회사명 입력으로 받아와 9번 기능과 같이 실행한다. |

3.12 하이퍼링크

| | |
|-------------|---|
| Actor | User, WebServer |
| Description | 입력한 회사의 잡플래닛 페이지로 바로 이동하는 아이콘을 배치하여, 사용자가 클릭 시 바로 이동할 수 있도록 한다. |

3.13 회사명 입력 기록 저장

| | |
|-------------|------------------------------------|
| Actor | User, DB |
| Description | 사용자의 아이디를 기반으로 입력되는 회사명을 db에 기록한다. |

3.14 입력어 자동완성 기능

| | |
|-------------|---|
| Actor | User, DB |
| Description | 잡플래닛 사이트에 존재하는 모든 회사명 목록을 db에 기록해두고, 회사명이 입력될 때, 자동 완성을 도와준다. |

4. Concept of operation

4.1 로그인

| | |
|----------|---|
| Purpose | 시스템을 사용하기 위해 등록된 사용자인지 확인 |
| Approach | 사용자가 시스템을 실행 후 로그인 시, ID, PW를 입력 후 로그인을 요청하면 웹서버에서 회원 정보를 검증 후 로그인 성공/실패 여부 확인한다. |
| Dynamics | 시스템 실행 시 로그인할 경우 |
| Goals | 로그인 기능을 구현한다. |

4.2 로그아웃

| | |
|----------|---|
| Purpose | 다른 사용자가 시스템을 사용하기 위해 로그인 상태를 초기 상태로 되돌린다. |
| Approach | 사용자가 시스템 실행 중 로그아웃 시, 로그인 상태를 초기 상태로 되돌리고, 로그인 화면으로 돌아간다. |
| Dynamics | 시스템 실행 중 로그아웃할 경우 |
| Goals | 로그아웃 기능을 구현한다. |

4.3 회사명 입력

| | |
|----------|--|
| Purpose | 시스템에 리뷰 요약 정보를 확인할 회사명 입력값 확인 |
| Approach | 사용자가 시스템을 ID, PW를 입력 후 검색을 요청하면 웹서버에서 회사명을 조회 후 회사명 존재 여부를 확인한다. |
| Dynamics | 시스템 실행 후 하나의 회사명을 입력하여 검색할 경우 |
| Goals | 회사명 입력에 대한 검색 기능을 구현한다. |

4.4 원하는 정보 항목 선택

| | |
|----------|---|
| Purpose | 리뷰 요약 정보를 출력하기 앞서 원하는 정보만을 출력하기 위해 항목을 선택 |
| Approach | 회사명을 검색하고 난 후, 작성날짜, 별점, 장점, 단점, 경영진에게 바라는 점 등 잡플래닛 리뷰에 있는 다양한 항목 중 원하는 항목을 선택한다. |
| Dynamics | 회사명 입력 후 리뷰 요약 이전에 세부 사항을 설정하는 경우 |
| Goals | 로그인 기능을 구현한다. |

4.5 리뷰 요약

| | |
|----------|--|
| Purpose | 사용자가 원하는 회사의 원하는 항목에 대한 리뷰 요약 정보 제공 |
| Approach | 회사명 입력 및 세부 항목을 선택 후, 해당 항목에 대해 잡플래닛 웹서버에서 크롤링을 거쳐 리뷰 요약 정보를 하나의 엑셀 파일로 요약하여 저장한다. |
| Dynamics | 리뷰 요약 정보를 출력하는 경우 |
| Goals | 리뷰 요약 파일을 생성 후 출력을 구현한다. |

4.6 면접 기출 정리

| | |
|----------|--|
| Purpose | 사용자가 원하는 회사의 면접 정보를 모아 정리한 문서 파일을 제공 |
| Approach | 앞서 사용자가 회사명 입력 후, 잡플래닛 웹서버에서 크롤링을 거쳐 면접 질문을 문서 파일로 정리해서 추출하여 저장한다. |
| Dynamics | 면접 기출집 생성 후 출력하는 경우 |
| Goals | 면접 기출집 생성 후 출력을 구현한다. |

4.7 파일 공유

| | |
|----------|--|
| Purpose | 사용자가 리뷰 요약 파일이나 면접 기출집 파일을 바로 공유하는 기능 제공 |
| Approach | 구글 계정과 보내는 이메일 주소를 입력하고 회원 정보를 검증 후 로그인 성공/실패 여부를 확인한다. 성공시 생성된 리뷰 요약 파일이나 면접 기출집 파일을 해당 이메일의 첨부파일로 포함시켜 전송한다. |
| Dynamics | 생성된 파일을 바로 공유하는 경우 |
| Goals | 공유 기능을 구현한다. |

4.8 여러 회사명 입력

| | |
|----------|--|
| Purpose | 시스템에서 여러 회사명 입력값 확인 |
| Approach | 기존에는 회사명 입력창이 하나인데, 추가 버튼을 통하여 입력창을 여러 개로 늘려 여러 회사명을 입력할 수 있게 한다. 사용자가 시스템을 ID, PW를 입력 후 검색을 요청하면 웹서버에서 입력된 여러 회사명을 조회 후 회사명 존재 여부를 모두 확인한다. |
| Dynamics | 시스템 실행 후 여러 회사명을 입력하여 검색할 경우 |
| Goals | 여러 회사명 입력에 대한 검색 기능을 구현한다. |

4.9 여러 회사의 리뷰 요약

| | |
|----------|--|
| Purpose | 사용자가 입력한 여러개의 회사의 원하는 항목에 대한 리뷰 요약 정보 제공 |
| Approach | 앞서 사용자가 여러 개의 회사명 입력 후, 세부 항목을 선택하고 나서, 해당 항목에 대해 각각의 회사 리뷰를 잡플래닛 웹서버에서 크롤링을 거쳐 수집한다. 그 후 입력된 모든 회사의 리뷰 요약 정보를 하나의 엑셀 파일로 요약하여 저장한다. |
| Dynamics | 여러 회사에 대한 리뷰 요약 정보를 출력하는 경우 |
| Goals | 여러 회사에 대한 리뷰 요약 파일을 생성 후 출력을 구현한다. |

4.10 팔로잉 목록 확인

| | |
|----------|--|
| Purpose | 사용자의 계정 정보 중 팔로잉 목록을 불러와 출력 |
| Approach | 팔로잉 목록을 확인하기를 요청하면 잡플래닛 웹서버에서 로그인 된 회원 정보를 이용해 조회 후 팔로잉 목록 출력한다. |
| Dynamics | 로그인 상태에서 팔로잉 목록 확인할 경우 |
| Goals | 팔로잉 목록 출력하는 기능을 구현한다. |

4.11 팔로잉 중인 회사들의 리뷰 요약

| | |
|----------|---|
| Purpose | 회사명 입력을 거치지 않고, 팔로잉된 기업끼리 비교 |
| Approach | 로그인 중인 사용자가 팔로잉 목록을 확인 후 팔로잉 목록을 검색할 여러 회사명들의 입력값으로 이용해 팔로잉 중인 모든 회사의 리뷰 요약 정보를 하나의 엑셀 파일로 요약하여 저장한다. |
| Dynamics | 팔로잉 중인 회사에 대한 리뷰 요약 정보를 출력하는 경우 |
| Goals | 팔로잉 중인 회사에 대한 리뷰 요약 정보를 출력하는 기능을 구현한다. |

4.12 하이퍼링크

| | |
|----------|---|
| Purpose | 리뷰를 출력하는 경우 해당 회사의 잡플래닛 주소로 바로 가는 버튼 제공 |
| Approach | 리뷰를 출력하는 경우 출력하는 모든 회사의 잡플래닛 주소로 바로 가는 버튼을 각각 제공한다. 이때, 각각의 버튼에 하이퍼링크 기능을 사용한다. |
| Dynamics | 앱 실행 시 로그인할 경우 |
| Goals | 로그인 기능을 구현한다. |

4.13 회사명 입력 기록 저장

| | |
|----------|--|
| Purpose | 사용자의 로그인 ID를 기반으로 DB에 입력 기록을 저장 |
| Approach | 사용자의 로그인 ID를 기반으로 DB에 해당 ID에서의 입력 기록을 저장하고, 회사명 입력 시 최근에 입력한 기록을 순서대로 보여준다. 이때, 입력 기록은 최대 10개로 제한한다. |
| Dynamics | 회사명을 입력하는 경우 |
| Goals | 검색 기록 저장 및 출력 기능을 구현한다. |

4.14 입력어 자동 완성 기능

| | |
|----------|--|
| Purpose | 입력 중인 회사명에 대해 자동 완성을 제시한다. |
| Approach | 잡플래닛에 등록된 회사명들을 모두 저장한 DB 파일을 기반으로 사용자가 회사명을 입력할 때, 자동 완성을 제시하여 편의성을 제공한다. |
| Dynamics | 회사명을 입력하는 경우 |
| Goals | 입력어 자동완성 기능을 구현한다. |

5. Problem statement

5.1 Overview

JobBoost는 잡플래닛 사이트의 리뷰 정보를 크롤링하여 수집하고, 수집된 정보를 사용자가 원하는 형태로 요약하여 제공하는 것이 주요 목적이다. 사용자들이 취업 정보를 얻는 과정에서 효율성과 편의성을 제공하는 것이 대표적인 장점이다.

5.2 Problem

JobBoost가 사용자들에게 제공하는 효율성과 편의성을 높이기 위해선 다음과 같은 문제점들에 직면할 수 있으며, 해결 방법은 다음과 같다.

5.2.1 problem #1 - 합법적으로 크롤링하기

크롤링이라는 것의 개념에 대해 정확히 짚고 넘어 가야 한다. 웹 크롤링(Web Crawling)이란 기본적으로 웹 상에 존재하는 데이터를 자동적으로 탐색하는 행위를 의미한다. 웹 크롤링은 웹 상에 존재하는 데이터를 내가 원하는 내용으로, 원하는 수량만큼 수집할 수 있다는 점에서 큰 장점을 지니고 있다. 그래서 많은 데이터 분석가나 데이터 과학자들이 웹 크롤링을 시도하고 있다. 하지만 웹 크롤링 수행이 문제가 되는 경우가 있습니다. 경우에 따라 웹 크롤링 수행이 불법으로 간주되는 경우가 있기 때문이다. 모든 케이스를 일일이 따져가며 특정 경우에 웹 크롤링이 불법인가 아닌가에 대해 따져보는 것은 어려운 일이다. 하지만 웹 크롤링이 문제가 되는 경우를 판가름하는 몇 가지의 기준이 있고 이를 정확히 인지한다면 법을 잘 지키면서도 윤리적으로 웹 크롤링을 이용할 수 있다.

흔히 알려진 웹 크롤링을 안전하게 하기 위한 방법으로 다음 두 가지 사항을 지켜야 한다.

- 일반적으로 수집한 데이터를 상업적으로 이용하지 않아야 한다.
- 웹 크롤링 행위가 상대방의 서버에 문제를 일으켜서는 안 된다.

5.2 problem #2 - 효율적인 DB 관리

회사명 입력어 기록 저장과 입력어 자동 완성 기능을 위해서는 DB의 사용이 필수적이다. 특히, 입력어 자동 완성 기능을 위해서는 2만개가 넘는 회사명을 DB에 저장하고, 이를 효율적으로 관리할 수 있어야 할 것이다. 회사명을 입력 도중 자동완성으로 회사명을 빠르게 제시하기 위해선 탐색 알고리즘의 응용이 필요할 것으로 보인다. 또한, 회사명 입력어 기록 저장 기능에서는 한 계정에 대한 기록에 사용되는 DB의 저장공간을 제한해야 한다. 이러한 과정에서 DB에 새로운 정보를 추가하고, DB가 가득 찬 경우에는 DB의 가장 오래된 내용을 삭제하는 복잡한 과정이 이루어 질 것이다. 결론적으로, DB를 완벽히 구현하기 위해서는 공부를 해야 할 것이다.

5.3 problem #3 - 요약 정보의 가독성 및 가시성

JobBoost는 수집된 정보를 사용자가 원하는 형태로 요약하여 제공한다. 요약된 정보를 출력할 때, 사용자가 읽기 쉽고, 한눈에 잘 보여야 하는 것이 중요하다. 리뷰 요약 정보 출력에 대해서는 사용자가 리뷰에서 원하는 세부 항목에 관해서만 출력된다. 그렇기에 출력되는 정보의 양식이 항상 일정하지 않을 것이다. 이 정보의 경우 엑셀 파일로 제공되며, 선택에 따라 항상 달라지는 열(세부항목)에 대해서 우선 순위를 주어 가시성을 높이고, 통일성을 갖추어야 한다. 반면 면접 기출 정보 출력의 경우 문서 파일로 제공되며, 양식을 설정하여 가독성을 보장해야 한다.

6. Glossary

| Term | Description |
|------|--|
| 취준생 | 취업준비생을 줄여 일컫는 말로, 일반적으로는 대학 고학년, 및 이미 졸업한 고학번 졸업생으로 구직 활동을 하는 학생 혹은 기존에 다니던 회사를 그만두고 새롭게 직장을 구하려 하는 사람 |
| 플랫폼 | 서비스가 운영되고 있는 환경 |
| 크롤링 | 소프트웨어와 같은 무언가가 인터넷을 돌아다니며 정보를 수집해 오는 작업 |
| 팔로잉 | 어떠한 것의 행동을 추적하는 것을 뜻한다. 팔로잉을 통해 쉽게 소식을 얻을 수 있다. |
| 리뷰 | 어떠한 것에 대해서 간단하게 설명이나 소개 |
| DB | 데이터베이스 / 여러 사람이 공유하여 사용할 목적으로 체계화해 통합, 관리하는 데이터의 집합 |
| 가독성 | 텍스트를 읽고 얼마나 쉽게 이를 이해할 수 있는가의 정도를 의미 |
| 가시성 | 눈에 얼마나 잘 보이는 정도를 이르는 말 |

7. References

잡플래닛

<https://www.jobplanet.co.kr/job>

합법적인 크롤링

<https://brunch.co.kr/@8d1b089f514b4d5/33>

1) [그림 1] 통계자료

<https://www.yna.co.kr/view/AKR20221018158200002>

2) 잡코리아 통계자료 인용

https://www.jobkorea.co.kr/goodjob/tip/view?News_No=19373&schCtgr=120001