

# 채팅 봇과 빅데이터를 활용한 온라인 매칭 시스템

\*함성호,\*정예진,\*김우영,\*박종민,\*오은주  
\*한성대학교 정보시스템공학과\*  
\*이화여자대학교 컴퓨터공학과\*  
\*건국대학교 소프트웨어학과\*  
\*강남대학교 컴퓨터공학과\*  
\*상명대학교 미디어소프트웨어학과\*

e-mail : [haams704@gmail.com](mailto:haams704@gmail.com)

Using Chat-bot and BigData to match the hospital service  
with person who needs to be cured.

Seong Ho Haam , Ye Jin Jeong , Woo Young Kim ,  
Jong Min Park , Eun Ju Oh

\*\*Dept of Information Engineering System , Hansung University  
\*\*Dept of Computer Engineering , Ewha woman's University  
\*\*Dept of Software Engineering , Kunkuk University  
\*\*Dept of Computer Engineering , Kangnam University  
\*\*Dept of MediaSoftware Engineering , Sangmyung University

## 요 약

클라우드 기반 채팅 봇 인프라와 Hadoop을 Mash-Up한 사용자 맞춤 병원 소개 서비스

### 기획 배경

바쁜 현대인들이 병원을 검색하기 위하여 인터넷을 켜는 수고로움을 덜고 자신에게 딱 맞는 병원을 소개받고자 기획함

### 1.1 클라우드 서버 (전체 적인 필수 인프라)

## 1. 서론

최근 4차 혁명의 물결 속에서 AI 기술과 빅데이터 기술이 각광받고 있습니다.

현대인들은 이미 모바일 플랫폼에는 많이 익숙해져 있는 상태이고, 현재 서비스 되고 있는 여러 IT 기술에도 빠르게 적응하는 세대입니다.

이렇게 급변하는 IT 트렌드 흐름 속에서 가장 중요한 가치를 선별한다면 자동화시스템, 클러스터링, 그리고 빠른 응답속도와 정확한 반환 가치라 할 수 있습니다.

사용자가 원하는 정보만을 필터링하여 빠르고 정확하게 전달하는 기술 (선별 기술 - 클라우드 인프라), 질문에 맞게 미리 구축된 구조대로 자동 응답을 하는 AI 기술, 들어온 콘텐츠를 의미 있는 데이터로 정리하고 이를 클러스터링 작업을 통해 더 구조화된 관리가 가능한 빅데이터, 이 세 가지 핵심 가치에 초점을 맞추어 프로젝트를 진행하게 되었습니다.

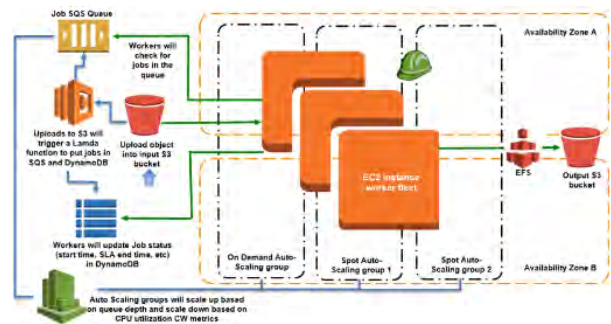


그림1. 클라우드 인프라 - AWS

전체적인 메시징 시스템을 관리하기 위해선 그림1 과 같은 시스템 인프라가 구축 되어야 합니다.

아마존 웹서비스에서 제공되는 클라우드 서버로 EC2를 사용하고 원격 SQL서버 RDS를 활용하여 메인 데이터베이스를 관리합니다. 또한 공공 데이터는 HDFS (Name-node)에서 관리하고 scoop 과정을 통하여 메인

데이터베이스 서버로 보내게 됩니다.  
보낸 데이터는 클라우드 서버에서 관리되고 사용자의 요구에 맞게 정제되어 제공되게 됩니다.

이러한 서비스가 제대로 제공하려면 클라우드 서버에서는 클라이언트의 요구사항을 관리하고 거기에 맞게 알고리즘이 구축되어 있어야 합니다.

본 프로젝트에서는 사용자들이 보내는 메시지를 순서대로 관리하고 사용자들이 필요로 하는 정보들을 흐름에 맞게 정제하여 보내주는 역할을 합니다.

```

root@ip-172-31-11-196: /home/ubuntu/hanium_chatbot_test/kakao_chatbot_test/haniu...
db is connected well
GET /keyboard 200 7.030 ms - 144
DBr8smWyVoo2
버튼클릭 인식 실패
{ user_key: 'DBr8smWyVoo2', type: 'text', content: '피부' }
POST /message 200 31.738 ms - 2
DBr8smWyVoo2
버튼클릭 인식 실패
{ user_key: 'DBr8smWyVoo2', type: 'text', content: '눈' }
POST /message 200 2.157 ms - 2
GET /keyboard 200 1.110 ms - 144
DELETE /chat_room/DBr8smWyVoo2 404 7.628 ms - 2167
GET /keyboard 200 0.718 ms - 144
DBr8smWyVoo2
새 번째 버튼 클릭
{ user_key: 'DBr8smWyVoo2',
  type: 'text',
  content: '가야할 곳에서 가까운 병원들' }
POST /message 200 1.480 ms - 94
DBr8smWyVoo2
버튼클릭 인식 실패
{ user_key: 'DBr8smWyVoo2', type: 'text', content: '서울' }
POST /message 200 1.353 ms - 94
    
```

그림 2. 서버 인프라 구성

1.2 채팅 봇 서비스 (인프라 내 구축된 API)



그림3. 채팅 봇 서비스 (Kakao.api활용)

채팅 봇 서비스는 이미 생활 속에 정말 많은 곳에서 자리 잡고 있습니다.  
가장 대표적인 채팅 봇 서비스 활용 분야는 SNS 서비스입니다.

그림3은 본 프로젝트에서 작업한 내용입니다.  
이와 같이 사용자가 원하는 정보들을 버튼 형식이나 이미지 또는 텍스트 형식으로 메시지를 전달받습니다.  
전달 받은 정보는 관리자가 관리하는 것이 아닌 서버 인프라 내에 구축된 채팅 봇이 미리 정리된 시스템적 흐름에 맞게 사용자에게 필요한 정보를 정제하여 제공하게 됩니다.

위와 같은 채팅 봇 서비스는 이미 여러 기업에서 프레임워크나 API를 제공하고 있습니다.  
대표적인 예로 그림4와 같이 msAzure , aws Lex , Watson ai 등이 많습니다. 이들은 주로 SNS에서의 api들과 접목하여 그 기능을 최대치로 발휘하고 있습니다.  
이러한 채팅 봇 서비스는 구축한 서버 인프라에 맞게 적용하여 활용하면 됩니다.



그림4. 채팅 봇 API 제공 서비스  
- msAzure bot Framework  
- Amazon Lex (with Lambda)  
- IBM Watson (conversation service)

또한 본 프로젝트에서는 채팅 봇 서비스 뿐만 아니라 이미지 딥러닝을 통한 정보 추출 작업까지 진행합니다.

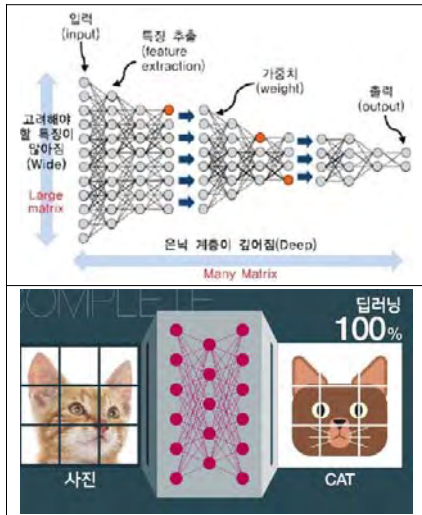


그림5. 이미지 딥러닝을 통한 정보 추출

아픈 부위를 버튼 형식으로 선택하여 정보 분류를 할 수 있으나, 다른 방법으로도 정보를 분류하여 제공하고자와 위와 같은 이미지 딥러닝 기술을 추가 하였습니다. 위 과정에서 TensorFlow 기술을 사용하여 이미지 분석과 학습을 진행하고 인식률을 높이기 위해 구글에서 제공하는 Cloud vision api도 추가로 사용하였습니다. 진행 과정에서 단순 정적 이미지를 분석하기 때문에 많은 시간이 소요되지 않아도 정확한 정보를 추출해내어 제공할 수 있었습니다. 이러한 기술들을 프로젝트에 접목 시켜 사용자들은 본인이 아픈 부위를 촬영하여 업로드 시에 어디가 아픈지 자동으로 체크하도록 하고 그에 맞는 병원 리스트를 뿌려주어 사용자가 필요한 정보만을 받아볼 수 있도록 진행 됩니다.

### 1.3 빅데이터 (공공 데이터 관리)

본 프로젝트에서는 빅데이터를 병원 공공 데이터 API를 크롤링하여 미리 구축한 원격 데이터베이스 서버(RDS)에 데이터를 넣어둡니다. 넣어둔 데이터는 scoop 과정을 통해 필요한 정보만을 가지고 와서 분산 처리 데이터베이스 서버에 저장합니다. 저장한 데이터는 다시 서버 인프라 내의 메인 데이터베이스(SQL서버)로 데이터들을 정제하여 보냅니다. 보내진 데이터는 사용자의 니즈에 맞게 데이터를 추가 정제하여 알맞게 제공됩니다.

## 본론

### 1. 주요 기술

```
GET / 304 510.584 ms --
GET /stylesheets/style.css 304 3.405 ms --
*강북삼성병원 / 서울특별시 종로구 새문안로 29 (평동) / 126.967812251
*건국대학교병원 / 서울특별시 광진구 능동로 120-1 (화양동) / 127.0720
*경희대학교병원 / 서울특별시 동대문구 경희대로 23 (회기동) / 127.0501
*고려대학교의과대학부속구로병원 / 서울특별시 구로구 구로동로 148 (구로
*삼성서울병원 / 서울특별시 강남구 일원로 81 (일원동, 삼성의료원) / 12
*서울대학교병원 / 서울특별시 종로구 대학로 101 (연건동) / 127.000398
*연세대학교의과대학 강남세브란스병원 / 서울특별시 강남구 언주로 211 (
*이화여자대학교의과대학부속목동병원 / 서울특별시 양천구 안양천로 1071
*재단법인아산사회복지재단 서울아산병원 / 서울특별시 송파구 올림픽로43길
*중앙대학교병원 / 서울특별시 동작구 흑석로 102 (흑석동) / 126.960419
*학교법인 고려중앙학원 고려대학교의과대학부속병원(안암병원) / 서울특별
*학교법인가톨릭학원가톨릭대학교서울성모병원 / 서울특별시 서초구 반포대
```

그림 6. 공공 데이터 SCOOP

**인프라 측면** : 4차 혁명의 핵심 AI기술과 빅데이터 , 그중 채팅 봇 시스템과 HDFS 소개

- > 사용자의 응답에 따라서 클라우드 서버에 구축한 채팅 봇 알고리즘에 맞게 플로우가 진행됩니다.
- > AWS devops 활용 / Node.js , MongoDB를 통한 기본 인프라 구축 / Kakao.api를 통해 채팅 봇 인프라 구축
- > Git / GitHub를 통한 형상 관리 및 이슈 관리

서울시 공공 데이터 부문에서 서울시 병원 API를 크롤링하여 Hadoop의 데이터노드인 HDFS에 저장합니다. 저장한 데이터들은 사용할 데이터들만 Map-Reduce를 통해 Scoop하고 이를 분산 처리된 DB에 저장합니다. 분산 처리된 DB에서는 기존 SQL-DB에서 SQL문을 활용하여 사용자가 필요한 정보만을 추출하여 주요 서비스로 제공됩니다.

**서비스 측면** : 사용자 입장에서 쉽게 접할 수 있는 플랫폼 활용

- > 채팅 봇 알고리즘과 빅데이터를 활용하여 정제된 데이터는 사용자들이 쉽게 접할 수 있는 대중화된 플랫폼인 카카오톡에서 서비스가 제공 됩니다.
- > Kakao.api 활용 à 서비스 플랫폼 구축
- > Restful.api 활용 à 시스템 아키텍처 구축

주요 기술로 1.3에서 설명한 Kakao.api를 활용하여 채팅 봇 서비스를 구현합니다. 카카오톡은 남녀노소 불문하고 가장 대중적인 SNS 플랫폼이라는 것은 바뀌지 않는 사실입니다. 따라서 본 프로젝트는 클라이언트 플랫폼을 카카오톡을 활용한 서비스이기 때문에 사용자들이 쉽게 접하고 거리감 없이 사용할 수 있을 것입니다.

## 2. 시스템 구조

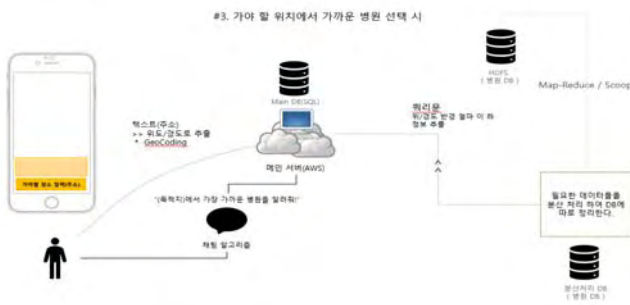


그림 7. 가야할 목적지에서 가까운 병원 검색 시

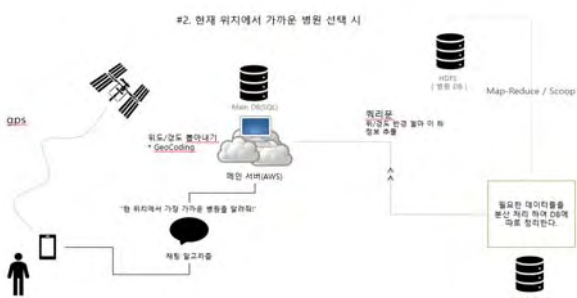


그림 8. 현재 위치에서 가까운 병원을 선택할 경우

### [빅 데이터]

- > 공공 데이터 API를 활용하여 데이터 수집하여 RDS에 내장 시키고, 분산 처리하기 전 모든 데이터들을 RDS에서 가지고 와서 HDFS에 관리합니다. .
- > 사용할 필요 데이터 Map-Reduce 작업을 하고 Scoop을 통해 데이터를 분산 처리를 합니다.
- > SQL 서버와 연동 하여 사용자에게 정말 필요한 데이터 들만 추출합니다.
- > 카카오 API를 활용하여 응답 형태로 서비스 제공됩니다.

### [채팅 봇 알고리즘 + 클러스터링]

- > 사용자가 원하는 정보를 텍스트로 관리합니다.  
(ex : 현재 위치 GPS정보 : 위도/경도로 뽑아 사용할 수 있게 정리)
- > 버튼이나 텍스트로 물어본 질문에 대해선 필요 데이터 들만 자연어 처리 (# - 해시태그 , @ - 어노테이션) 됩니다.
- > 키워드를 기반으로 어떠한 응답을 내줘야 하는지 선 클러스터링 작업을 통해 Hadoop의 Name-Node에 데이터 요청을 합니다.
- > 요청된 데이터는 후 클러스터링 작업을 통하여 필요 데이터만 추출하고 SQL 문을 활용하여 사용자에게 서비스가 제공됩니다.

## 결론

### ㄱ. 기대효과

인공지능 봇을 활용한 서비스는 앞으로도 많이 나올 것이고, 데이터는 계속 해서 차고 넘칠 것입니다.

이렇게 급변하는 IT 흐름 속에서 봇과 빅데이터를 활용한 서비스는 필수 불가결한 존재가 되었습니다.

이러한 핵심 기술들을 사용하여 여러 분야에 활용한다면 사용자가 정말 원하는 정보만을 빠르고 정확하게 뽑아내고 이를 자동화 시스템을 통하여 편하게 제공할 수 있을 것입니다.

### ㄴ. 활용방안

빅데이터와 채팅 봇을 활용한 서비스는 위에 활용 트렌드 비교 부분에서도 볼 수 있듯이, 상당히 많은 곳에서 활용되고 있습니다.

여기에 덧붙여 본 프로젝트와 같이 대중화된 플랫폼과 연결 지어 서비스 한다면 그 과급력과 사회 기여도는 어마어마할 것 같습니다.

예를 들어, 카카오톡 내에 카카오 Navi, 카카오 Bank 등 여러 서비스에 위와 같은 인프라를 구축한다면, 사용자들을 독실할 만한 강력한 서비스 플랫폼이 완성될 것이라 생각합니다.

## Reference

- <http://www.bizwatch.co.kr/pages/view.php?uid=8066>
- \*\* 빅데이터 활용 분야 - 마케팅  
<http://www.sedaily.com/NewsView/1ODK4QW240/GD02>
- \*\* 채팅 봇 활용 분야 - 백화점  
<https://ap-northeast-2.console.aws.amazon.com/ec2/v2/home?region=ap-northeast-2#>
- \*\* 본 프로젝트의 EC2 사이트 (로그인 해야 입장 가능) (server infrastructure)  
<https://www.ibm.com/watson/>
- \*\* IBM Watson ( Chatbot FrameWork)  
<https://center-pf.kakao.com/>
- \*\* 카카오 채팅 봇 (플러스친구)  
<https://aws.amazon.com/ko/rekognition/>
- \*\* 아마존 이미지 분석 딥러닝  
<https://www.kadenze.com/courses/creative-applications-of-deep-learning-with-tensorflow/info>
- \*\* TensorFlow 동영상 강좌  
<https://www.youtube.com/watch?v=a74pFg8paVc&list=PL1H8jIvbSoIqIXVcdZTH2xsYFp3e1Nmjf>
- \*\* TensorFlow Youtube 강좌