

카카오톡 API를 이용한 Django 기반 챗봇 시스템

고흥찬, 김민수, 이솔비, 이형우*
한신대학교 IT대학 컴퓨터공학부

Django based ChatBot System Using KakaoTalk API

Heungchan Ko, Minsu Kim, Solbi Lee, Hyung-Woo Lee*

Division of Computer Engineering, Hanshin University

요약 본 논문은 대학생들이 소속 대학 내 주요 정보를 손쉽게 빠르게 검색할 수 있도록 KakaoTalk API를 활용하여 장고 프레임워크를 이용하여 챗봇 시스템을 개발하였다. 특정 정보만을 제공하는 기존 챗봇 시스템과는 달리 본 연구에서 개발한 챗봇은 대학 내 주요 정보 뿐만아니라 날씨, 유튜브, 네이버 실시간 랭킹 검색 및 언어 번역 등 다양한 형태의 사용자 질의에 대해 자동으로 검색 결과를 제공한다. AWS Ubuntu 서버내에 Apache, Python 및 Django를 이용하여 모듈을 개발하였고 KakaoTalk API 및 BeautifulSoup 등을 이용하여 카카오톡 서버와 통신하여 사용자 질의에 대해 자동으로 응답하는 챗봇 시스템을 개발하였다. 본 연구에서 개발한 시스템에 대해서는 향후 대학 입시 정보 홍보 및 선거 홍보 시스템 등에 적용할 수 있을 것으로 기대된다.

주제어 : 챗봇, 카카오톡 API, Django, Python, 데이터 통신

Abstract In this paper, we developed a chatbot system using the Django framework using the KakaoTalk API so that college students can easily search for important information in their university. Unlike existing chatbot systems that provide only specific information, the chatbot developed in this research automatically provides search results for various types of user queries such as weather, YouTube, Naver real-time ranking search and language translation as well as important information within their own university. We developed a module using Apache, Python and Django in AWS Ubuntu server and developed a chatbot system that automatically responds to user queries by communicating with KakaoTalk server using KakaoTalk API and BeautifulSoup. The system developed in this study is expected to be applicable to the future university entrance information promotion and election promotion system.

Key Words : ChatBot, Kakao Talk API, Django, Python, Data transmission

1. 서론

현대 사회는 수많은 정보들로 넘쳐나기 때문에 이를 빠르고 정확하게 이용하는 것은 중요하다. 대학 내 학교 관련 주요 정보를 쉽게 얻지 못하는 새내기 또는 기존 학

생들이 많다. 학교 홈페이지나 알리미가 있지만 원하는 정보를 찾기에는 불편하고 시간도 많이 소요된다. 국내 주요 대학에서 제공하는 기존 챗봇 시스템을 조사한 결과 대다수 시스템이 간단한 형태의 정적 정보만을 제공하고 있으며, 도서관 열람실 자리 정보나 교통 정보 제공

*교신저자 : 이형우(hyungwoo8299@gmail.com)

접수일 2018년 04월 12일 수정일 2018년 05월 10일 심사완료일 2018년 5월 20일

등이 주요 기능이었다. 이에 대학 내 주요 정보에 대한 검색 결과를 제공하면서 사용자가 원하는 정보 등을 보다 손쉽게 빠르게 자동 제공할 필요가 있다.

이에 본 논문에서는 가장 많은 사용자를 확보하고 있는 카카오톡 서비스를 대상으로 KakaoTalk API를 활용하여 장고 프레임워크 기반의 챗봇 시스템을 개발하였다. 개발한 챗봇 시스템은 AWS Ubuntu 서버를 이용하였고, Python, Django, apache로 작업 환경을 구축하였고, KakaoTalk API, Sqlite3, BeautifulSoup를 활용하여 카카오톡 서버와 통신하여 사용자에게 다양한 정보를 제공할 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 챗봇에 사용된 관련 기술들에 대하여 서술하였고 3장에서는 시스템 설계에 대한 전체적인 시스템 구조, 서버 구조, 크롤링 동작 구조, DB 구조, 시스템 기능을 기술하고 있다. 4장에서는 시스템 구현 결과에 대해 설명하였으며 5장에서는 결론 및 향후 연구 내용을 제시하였다.

2. 관련기술

2.1 Apache HTTP Server 및 JSON

기존 카카오톡 서버와 연동하기 위해서는 검색 결과에 대해 연동 기능을 제공하기 위해 다양한 모듈을 사용해야 하난. 본 연구에서는 아파치 HTTP 서버[1]를 토대로 JSON[2] 모듈을 기반으로 개발하였다. JSON은 속성-값 쌍으로 이루어진 데이터 오브젝트를 전달하기 위해 인간이 읽을 수 있는 텍스트를 사용하는 개방형 표준 포맷으로, 특히, 인터넷에서 데이터를 송수신하는 과정에 적용 가능하다.

2.2 Django, Python 및 SQLite

Django[3]는 파이썬으로 작성된 오픈 소스 웹 애플리케이션 프레임워크로 모델-뷰-컨트롤러 패턴을 적용할 수 있다. Django를 사용하여 데이터베이스 기반 웹 사이트를 효율적으로 개발할 수 있으며, 콤포넌트의 재사용성과 플러그인화 가능성, 빠른 개발 등의 기능을 제공한다. 설정 파일부터 데이터 모델에까지 Python[4] 언어가 사용되었다. Python은 동적 범용 프로그래밍 언어로, 다양한 플랫폼을 지원하며, 라이브러리(모듈)가 풍부하여 대학을 비롯한 여러 교육 기관, 연구 기관 및 산업계에서

이용이 급증하고 있다. 또한, 머신러닝 기법 등에 적용 가능하여 챗봇 개발에 중심이 되는 언어로 선택하였다. 또한 DB 연동을 위해 SQLite[5]를 사용하였다.

SQLite는 MySQL나 PostgreSQL와 같은 데이터베이스 관리 시스템이지만, 서버가 아니라 응용 프로그램에 넣어 사용하는 비교적 가벼운 데이터베이스이다. 일반적인 RDBMS에 비해 대규모 작업에는 적합하지 않지만, 중소 규모라면 속도에 손색이 없다. 또 API는 단순히 라이브러리를 호출하는 것만 있으며, 데이터를 저장하는데 하나의 파일만을 사용하는 것이 특징이다.

2.3 Web crawler 및 BeautifulSoup

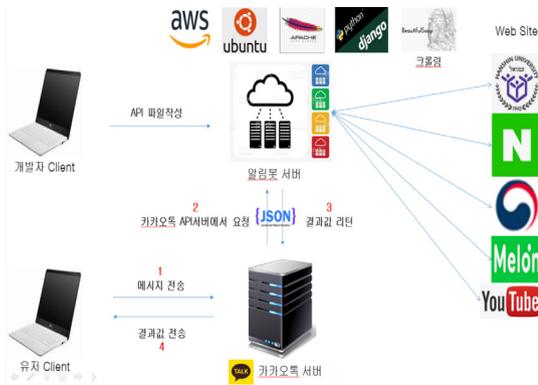
검색 엔진과 같은 여러 사이트에서는 데이터의 최신 상태 유지를 위해 웹 크롤링[6] 기법을 적용하였다. 웹 크롤러는 대체로 방문한 사이트의 모든 페이지의 복사본을 생성하는 데 사용되며, 검색 엔진은 이렇게 생성된 페이지를 보다 빠른 검색을 위해 인덱싱 한다. 또한, BeautifulSoup[7]을 이용하여 HTML parser 기능을 수행토록 하였다. 뷰티풀 수프는 HTML 및 XML 파일에서 데이터를 가져 오는 Python 라이브러리로 파서 트리를 탐색, 검색 및 수정하는 관용적인 방법을 제공하기 위해 사용하였다. 이밖에도, Python 언어로 XML과 HTML을 처리하기 위해 lxml[8]을 이용하였고, 또한 C 라이브러리 libxml2 및 libxslt에 대한 파이썬 바인딩 과정에 활용하였다.

3. 챗봇 시스템 설계

3.1 시스템 구조

본 연구에서 개발한 시스템의 전체 구조는 다음과 같다. 우선 사용자가 카카오톡 서버로 버튼이나 텍스트 형식의 메시지를 전송하면, 카카오톡 서버에서는 json 형식의 데이터를 챗봇 서버로 전달한다. 수신된 정보를 토대로 챗봇 서버에서는 Django를 기반으로 json 형식의 데이터를 Python 모듈을 이용하여 분석하고, 다시 json 형식의 검색 결과값을 반환해 주도록 하였다. 또한, 알림봇 서버에서는 사용자가 원하는 결과값을 카카오톡 서버로 반환해주고 마지막으로 카카오톡 서버는 사용자에게 해당 정보를 제공하게 된다. 사용자가 원하는 정보가 DB에 등록된 정보라면 바로 결과값을 반환하고 크롤링이 필요한 정보라면 각각의 웹 사이트들에서 얻은 정보를 반환

하게 된다. 이 때 챗봇 서버에 BeautifulSoup 라이브러리를 이용하여 html 형태로 작성된 코드를 Python 모듈이 분석하여 처리하며 파싱 결과를 화면에 다시 출력하도록 설계 및 구현하였다.



[Fig. 1] System Architecture

3.2 서버 구조

웹 서버는 웹 클라이언트가 특정 위치에 있는 서버 저장소에 있는 자원을 요청(HTTP request) 받아서 제공하는 응답(HTTP response) 처리가 기본 동작이다. 웹 애플리케이션 서버인 uWSGI는 웹 서버인 Apache와 웹 애플리케이션인 Django 간의 연결을 중계하는 역할을 한다. Apache는 파이썬을 모르기 때문에 uWSGI는 HTTP 요청을 python으로, Django로 부터 받은 응답을 Apache가 알 수 있도록 변환해준다.



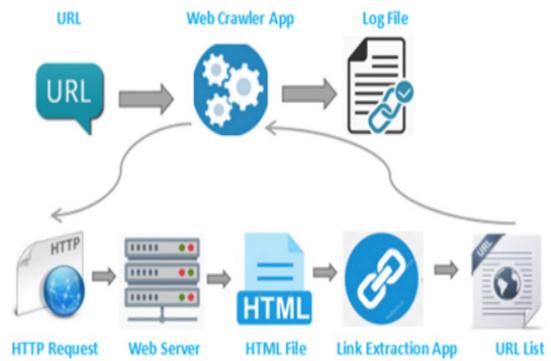
[Fig. 2] ChatBot System Interrelation Diagram

3.3 크롤링 동작 구조

크롤링을 처음 시작할 때 반드시 하나의 특정 웹 페이지의 URL 주소를 명시해주어야 한다. 보통 “Seed URL”라고 칭하는데 크롤링 프로그램이 가장 먼저 방문하는 URL로 중요한 역할을 한다. 왜냐하면 뒤이어 크롤링 되

어야할 웹 페이지의 URL은 “Seed URL” 웹 페이지가 링크하고 있는 웹 페이지를 순차적으로 방문하기 때문이다. 다시 말해서 초기 설정한 페이지를 방문하고 페이지가 링크하고 있는 모든 웹 페이지 URL을 수집하게 된다.

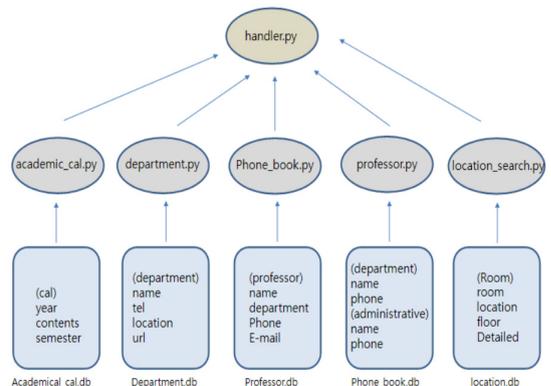
이후에는 수집한 HTML 페이지를 파싱한다. 파싱은 HTML의 태그와 유의미하지 않은 텍스트가 섞인 문서처럼 가공되지 않은 데이터에서 원하는 특정한 문자열을 빼내는 작업이다. 본 연구에서는 객체 파싱을 위해 파이썬 뷰티풀솅 라이브러리를 이용했다. 마지막으로 새로 발견된 URL을 순차적으로 방문하여 앞에 과정을 계속 반복 수행하게 된다.



[Fig. 3] Crawling Sequence Diagram

3.4 챗봇 Python 모듈 및 DB 구조

대학내 주요 정보를 제공하기 위해 학사일정, 부서, 전화번호, 교수 연구실 및 대학내 주요 시설에 대한 정보를 제공할 수 있도록 DB를 구성하였다. SQLite를 이용하여 DB를 관리하였는데 DB 등록시에는 gui 환경에서 DB Browser for SQLite를 이용해 작업이 용이했다.



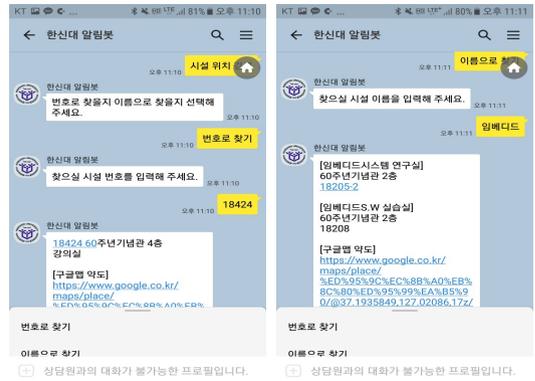
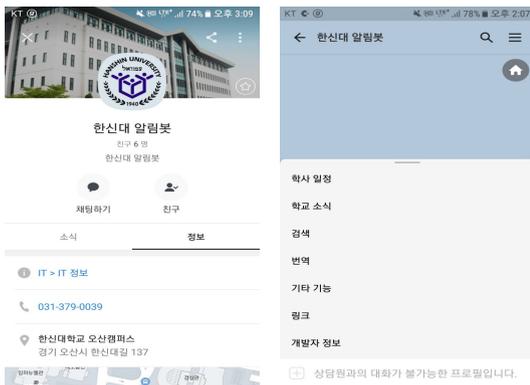
[Fig. 4] ChatBot System Program & DB Relationship

챗봇의 주요 기능은 다음과 같다. 검색 기능인 경우 전화번호, 학과, 교수, 시설 위치, 날씨, 유튜브, 뉴스 정보를 자동으로 제공하도록 구현하였다. 날씨 검색은 네이버 사이트에서 제공하는 정보를 크롤링하여 제공하며, 유튜브 사이트에서 검색 결과를 제공하도록 구현하였다. 뉴스는 네이버 뉴스 API를 이용했다. 또한, 언어 번역 기능을 제공하여 한국어, 영어, 중국어, 스페인어, 프랑스어까지 입력된 단어/문장에 대한 번역 결과를 제공하며 네이버 Papago NMT API를 이용했다.

이밖에도 기타기능으로는 네이버 실시간 순위, 네이버 영화 순위, 멜론 음악 순위, 도시별 현재 날씨 정보를 제공토록 하였다. 실시간 순위와 영화 순위는 네이버 사이트에서, 멜론 음악순위는 멜론 사이트에서, 도시별 현재 날씨는 기상청 사이트에서 크롤링 했다. 앞선 날씨 검색 기능과는 다르게 전국 도시별 날씨를 한눈에 볼 수 있도록 개발하였다.

4. 시스템 구현 결과

챗봇 구현 결과는 다음과 같이 기존 카카오톡에서 챗봇을 친구 추가하는 과정을 통해 손쉽게 본 연구에서 개발한 챗봇을 바로 이용할 수 있다. 친구 추가 과정 이후에 챗봇을 실행하면 아래 그림과 같이 주요 기능을 선택할 수 있도록 개발하였다. 그리고 '시설 위치 찾기' 기능을 선택한 후 해당 위치를 입력('18424')하면 위치 정보를 자동으로 보여준다.



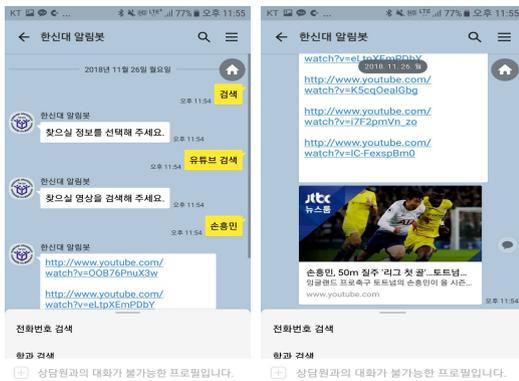
[Fig. 5] ChatBot Auto-Response System (Plus Friend & Query Auto-Response Results)

대학내 주요 소식에 대한 정보를 제공하며, 입력된 검색어에 대해 직접 대학내 웹사이트를 자동 크롤링하여 공지사항, 행사공지, 학사공지, 장학공지 및 취업공지 정보 등을 제공토록 하였다. 또한 부서 및 학부 위치와 전화 번호 등을 제공한다. 대학내 관련 학부 정보를 DB에 저장해 학과 전화번호, 시설위치 정보, 학과별 홈페이지 정보 등을 제공토록 하였다. 또한, 번호나 이름으로 시설 위치를 검색할 수 있으며 구글맵 API를 이용해 건물 사진도 제공하도록 구현하였다.



[Fig. 6] Weather Forecast Auto-Response Result

위 그림과 같이 주어진 전체 지역 중에서 원하는 지역의 날씨를 검색 할 수 있고 오늘과 내일의 날씨를 제공하며, 네이버 뉴스 API를 이용해 최신 뉴스(10개)를 자동으로 제공토록 구현하였으며, 아래 그림과 같이 유튜브 사이트 크롤링 후 5개의 영상을 자동 제공하였다.



[Fig. 7] YouTube Query & Result

또한, 네이버 Papago NMT API를 이용해 사용자가 입력한 단어 또는 문장에 대해 한국어, 영어, 중국어, 프랑스어, 스페인어, 프랑스어 번역 결과를 자동으로 제공토록 하였으며, 네이버 사이트를 크롤링 해 실시간 검색어는 20위, 실시간 영화 순위는 50위까지 정보를 제공토록 하였다. 그리고, 멜론 사이트를 크롤링 해 실시간 멜론 음악 순위 100위까지 제공하며 기상청 사이트를 크롤링 해 전국 도시별 날씨를 한눈에 볼 수 있도록 제공한다.



[Fig. 8] Auto Translation and Films and Music Real-Time Ranking Query Results

이와 같이 본 연구에서 개발한 카카오톡 API를 이용한 Django 기반 챗봇 시스템은 SNS 분야에서 가장 많은 이용자를 확보하고 있는 카카오톡 채팅 프로그램을 이용하여 사용자가 입력한 다양한 형태의 질의에 대해 검색 결과를 자동 응답하는 챗봇 프로그램을 개발하였다.

5. 결론

대학내 주요 정보는 대학내 홈페이지나 알리미 서비스 등을 통해 제공되는데 직접 검색 과정을 수행해야 하며, 원하는 정보를 빠르게 얻기에 불편하다는 단점이 있다. 카카오톡 채팅 시스템은 스마트폰 및 PC 상에서 손쉽게 대화 기능을 제공하고 있다. 따라서 최근에는 채팅 시스템 내에 자동화된 챗봇(ChatBot)이 개발되어 입력된 질의어에 대해 자동 응답하는 기능을 제공하고 있다.

이에 본 논문에서는 일반인 등에게 대학내 주요 정보에 대한 검색 결과를 자동으로 제공하고 또한 대학생들이 학교 관련 정보를 손쉽게 빠르게 검색할 수 있도록 KakaoTalk API를 활용하여 장고 프레임워크 기반 챗봇 시스템을 개발하였다.

본 시스템을 이용할 경우 학교 관련 정보를 빠르고 정확하게 제공하므로 검색에 소요되는 시간과 노력을 줄일 수 있으며, 또한 서비스 이용에 필요한 모듈이나 소프트웨어에 대한 추가 설치가 필요 없도록 구현하였다. KakaoTalk 인터페이스를 통해 자동응답 방식으로 구동하며, 사용자에게 친숙한 조작 방식으로 카카오톡 플러스 친구 추가 과정만 진행하면 바로 이용할 수 있다.

구현된 결과를 토대로 앞으로 AI 기능을 보다 개선하여 지능화된 챗봇으로 발전시키고자 한다. End-To-End 오픈소스 머신 러닝 시스템인 텐서플로우 모듈을 이용하여 현재 구현된 파이썬 모듈과 연동하여 사용자가 입력한 질의어에 대해 보다 지능화 응답 결과를 제공할 수 있도록 구현할 예정이다.

REFERENCES

- [1] Wikipedia “Apache HTTP Server”
https://en.wikipedia.org/wiki/Apache_HTTP_Server
- [2] Wikipedia “JSON”, <https://en.wikipedia.org/wiki/JSON>
- [3] Wikipedia “Django”,
[https://en.wikipedia.org/wiki/Django_\(web_framework\)](https://en.wikipedia.org/wiki/Django_(web_framework))

- [4] Wikipedia “Python”, [https://en.wikipedia.org/wiki/Python_\(programming_language\)](https://en.wikipedia.org/wiki/Python_(programming_language))
- [5] Wikipedia “SQLite”, <https://en.wikipedia.org/wiki/SQLite>
- [6] Wikipedia “Web crawler”, https://en.wikipedia.org/wiki/Web_crawler
- [7] Wikipedia “Beautiful Soup (HTML parser)”, [https://en.wikipedia.org/wiki/Beautiful_Soup_\(HTML_parser\)](https://en.wikipedia.org/wiki/Beautiful_Soup_(HTML_parser))
- [8] lxml <https://lxml.de>
- [9] Wikipedia “WSGI”, https://en.wikipedia.org/wiki/Web_Server_Gateway_Interface
- [10] Plus friend auto responder API, https://github.com/plusfriend/auto_reply
- [11] Papago NMT API, <https://developers.naver.com/docs/nmt/reference>
- [12] Naver News Search API <https://developers.naver.com/docs/search/news>
- [13] Park, Eung Yong, “Do it! Jump to Python”, Aegis Publishing, 2016
- [14] Tensorflow, <https://www.tensorflow.org>
- [15] Wikipedia “Data transmission”, https://en.wikipedia.org/wiki/Data_transmission

고 흥 찬(Heungchan Ko) [정회원]



- 2010년 2월 : 중흥고등학교 졸업
- 2012년 3월 ~ 현재 : 한신대학교 컴퓨터학과 재학중

<관심분야>

안드로이드 시스템, 정보보호, AI 시스템

김 민 수(Minsu Kim) [정회원]



- 2011년 2월 : 관악고등학교 졸업
- 2015년 3월 ~ 현재 : 한신대학교 컴퓨터학과 재학중

<관심분야>

안드로이드 시스템, 정보보호, AI 시스템

이 슬 비(Solbe Lee) [정회원]



- 2013년 2월 : 인덕원고등학교 졸업
- 2014년 3월 ~ 현재 : 한신대학교 컴퓨터학과 재학중

<관심분야>

안드로이드 시스템, 정보보호, AI 시스템

이 형 우(Hyung-Woo Lee) [종신회원]



- 1994년 2월 : 고려대학교 컴퓨터학과 (학사)
- 1996년 2월 : 고려대학교 컴퓨터학과 (석사)
- 1999년 2월 : 고려대학교 컴퓨터학과 (박사)
- 2003년 3월 ~ 현재 : 한신대학교 컴퓨터공학부 교수

<관심분야>

사물인터넷, 정보보호, 모바일 보안 및 디지털 포렌식 등