**What’s News**

**(Developer : 이승휘, 안형주, 곽민제)**

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Description | Editor | Version |
| 2018.10.10 | Server Controller 구성 업데이트 | 안형주 | V.0.1 |
| 2018.10.16 | Server Web Crawler 구성 부분 업데이트 | 안형주 | V.0.2 |

Index

[1 개요 3](#_Toc526428764)

[2 범위 4](#_Toc526428765)

[**2-1 Controller** 4](#_Toc526428766)

[**2-2 Web crawler** 4](#_Toc526428767)

[**2-3 Data base** 4](#_Toc526428768)

[3 구성 5](#_Toc526428769)

[**3-1 Controller** 5](#_Toc526428770)

[**3-2 Web crawler** 5](#_Toc526428771)

[**3-3 Data base** 5](#_Toc526428772)

1. ****개요****

What’s news는 Naver에서 제공하는 news를 script할 수 있도록 사용자에게 제공해주는 어플리케이션이다.

어플리케이션으로 제공해주는 news들은 Naver에서 제공하는 모든 분야의 news를 crawling 하여 사용자의 취향에 맞게 일정시간마다 업데이트 하여 제공해주는 것을 기본적인 작동원리로 한다.

사용자에게 제공되는 news들은 위젯의 형식으로 제공하며, Naver에서 등록되는 모든 기사를 단순히 제공해주는 것이 아닌, 동일한 주제의 기사는 filtering하여 전달해준다.

사용자는 news를 일정시간이 되면 server에 새로운 news의 전달을 request하고 server는 새롭게 업데이트한 news를 DB에서 확인하고 response 해준다. 이 과정에서 어플리케이션을 사용하든 모든 사용자가 동일한 시간에 server에 새로운 news를 요청하기 때문에 server의 과부하를 해결하기 위해 Rabbit MQ를 사용하여 client의 connection을 scheduling으로 관리해준다.

Server는 사용자의 IP 및 data를 관리하는 user database와 news를 저장 및 관리하는 news database를 접근하여 사용자의 request에 맞추어 database를 searching 하고 response 해주는 것이 기본적인 구조로 한다.

또한, news의 data를 filter 및 가공하기 위해 자연어처리 기반의 data가공의 작업이 이루어지며,

(data 가공 part 추후 업데이트 할 것.)

(data mining part도 추후 업데이트.)

1. ****범위****

사용자가 어플리케이션을 사용하면서 서버에 접속을 하고 news를 요청해야 하는 과정에서 Server가 어떻게 작동을 하는 것에 대해 기술한다. Server가 어떤 module로 구성되었으며, module들의 작동 원리에 대한 설명을 참조한다.

**2-1 Controller**

Controller는 Server의 전반적인 관리를 해주는 역할을 한다. 사용자와 server의 socket connection을 우선적으로 관리하며, 사용자의 request에 맞는 작업을 database를 참조하여 response 해주는 역할을 한다.

**2-2 Web crawler**

A

**2-3 Data base**

QGIS에서 검수를 실행할 수 있는 인터페이스에 관해 설명한다. 배치파일은 JAVA기반으로 개발되어졌으며, 인자값을 전달받아 실행된다. 호출방법 및 인자값등 실행하기 위한 모든 인터페이스에 대해서 설명한다.

1. ****구성****

Server는 java기반의 Controller, Web crawler, DB세 모듈로 나누어서 개발하고 각 모듈을 Server.Controller의 main에서 참조하도록 구현하는 것을 목적으로 하고 있다.

또한 lombok.jar, json-simple.jar, gson.jar 라이브러리를 참조한다.

**3-1 Controller**

Controller는 Server의 전반적인 관리를 하는 매니저 역할을 한다. Main, Connection, Processing 세가지 모듈과 News, Request, Response객체들로 구성되어져 있는 Message 패키지가 존재하며 Main은 Controller를 중앙에서 관리하는 역할을, Connection은 클라이언트와의 Connection을, Processing은 클라이언트의 Request를 처리하여 Response해주는 역할이다.

세부적으로는 Main에서 Connection의 connect메소드를 실행하여 스레드로 new client의 커넥션을 관리하고 Server socket을 생성해준다. 그리고Connection 스레드의 run을 통해 다시 Main의 deliverToProcess 메소드를 실행하여 Processing의 getRequest 메소드를 실행, 파라미터로 받은 Requset객체를 새로운 Processing 스레드로 관리하고, Gson으로 Request를 처리한다.

**3-2 Web crawler**

먼저 네이버 뉴스를 crawling하기 위해서 네이버 뉴스의 섹션 분류 ID를 분석 후, 2개의 HashMap(sid1: 대분류, sid2: 소분류)으로 분류했다. HashMap의 key와 value는 string으로 구성하였는데, 이는 섹션의 sid값이 변경될 경우 발생할 수 있는 섹션 중복문제를 방지하기 위함이다.

**3-3 Data base**

Python으로 Java 배치파일(jar)을 연동 및 호출한다.