Contents

1. Introduction to API ........................................................................................................................................2
   1.1. Sign-up for an API Key .........................................................................................................................2
   1.2. Forming a Request ................................................................................................................................8
2. Using Java to do data scraping ..................................................................................................................9
   2.1. The ApiExample ...................................................................................................................................9
   2.2. Coding a java file ..................................................................................................................................13
      2.2.1. Replacing the apiKey ....................................................................................................................14
      2.2.2. Change the xml page you want to retrieve ....................................................................................15
      2.2.3. Change the data you want to retrieve .............................................................................................15
      2.2.4. Print out the data you want ...........................................................................................................17
   2.3. Compile and execute your java file .....................................................................................................17
3. Load data into MySQL database ...............................................................................................................18
1. Introduction to API
Web API is typically defined as a set of Hypertext Transfer Protocol (HTTP) request messages, along with a definition of the structure of response messages, which is usually in an Extensible Markup Language (XML) or JavaScript Object Notation (JSON) format. The Ohloh (Openhub) API is a free, REST-based programming interface to the Ohloh (Openhub) open source directory. You can use the Ohloh API to create your own applications and web services based on Ohloh data.

You can get the detail documents of ohloh API from

https://github.com/blackducksw/ohloh_api#ohloh-api-documentation

1.1. Sign-up for an API Key
Before you can access the Ohloh (Openhub), you must register your application and obtain an API key. Bandwidth will initially be limited to 1,000 requests per API key per day.

https://www.openhub.net/

click “Join Now”
Activate your account from your email account!

Click Settings in the webpage
Click **API Keys**
Click Request New API Keys
After signing up, you can get your api key from your account information as follows:
Welcome to the Open Hub API program! Your new API key is shown below.

Settings: API Keys

networktutorial for the tutorial

API Key: 9Df4Zc5nF0d2xXNiIA
OAuth Secret: t9NtC0xKc2J1KqOxM6kKCC0f3sDwpB1GkK4t10

Created less than a minute ago. Hasn’t been used yet
Active: ✔ can be used
0 requests all-time
0 requests today
Limited to 1,000 requests a day

Edit API Key

About API Keys

- API keys are associated with your account, and a particular application or usage. Please refer to the Open Hub Terms of Use and API Agreement for more information.
- On this page, you can request new API keys, or you can edit or remove existing API keys. You can also see your usage statistics for the API keys you have been granted.
- For more information on the Open Hub API, please see the documentation.

Account Summary

Settings

Contributions

Recognition

Usage

Website Edits
1.2. Forming a Request
The Ohloh API returns XML-formatted data in response to HTTP GET requests.

Each web page on Ohloh, there may be an equivalent XML-formatted version of the page.

You must do three things to receive an XML-formatted response:

Append a .xml extension to the basic URL. For example, instead of http://www.ohloh.net/projects/1, which returns an HTML page, you would request http://www.ohloh.net/projects/1.xml.

Provide your API Key as an HTTP parameter. Your request will be forbidden without a valid api_key.

For example, to view the accounts in page 1 as XML, using an example API key, the complete URL would be:

https://www.ohloh.net/accounts.xml?api_key=wgXz1VXnmYC0PuALXQGPA&page=1

you can get a page like this:
2. Using Java to do data scraping

2.1. The ApiExample
Ohloh offers a tutorial on how to use api via an example. In this section, we will use this example to introduce some important codes.

You can download the example from

http://meta.ohloh.net/wp-content/uploads/2012/02/ApiExample.java

This example is used to retrieve data from the url

```java
URL url = new URL("http://www.ohloh.net/accounts/" + emailDigest + ".xml?api_key=" + apiKey + "&v=1");
```

Where emailDigest is the message digest of the account’s email address

Take my account as an example, the url would be

https://www.ohloh.net/accounts/230570.xml?api_key=wgXz1VXnmYCOPUalXQGPA&v=1
URLConnection con = url.openConnection();
DocumentBuilder builder = DocumentBuilderFactory.newInstance().newDocumentBuilder();
Document doc = builder.parse(con.getInputStream());

Is to create a document from this URL’s input stream and parse.

NodeList responseNodes = doc.getElementsByTagName("response");
Is to get the elements between the tags of “response”, as shown in the following picture.
Element resultElement = (Element)element.getElementsByTagName("result").item(0);

Is to get the elements between the tags of “result”
String realName = accountElement.getElementsByTagName("name").item(0).getTextContent();
System.out.println("Located the real name: " + realName);

Is to get elements between the tags of “name” and print out the result.
2.2. Coding a java file

You can use the java file we provided and have some revisions depending on your scraping needs.
In the following, we will take Ohlohaccounts.java as an example.

2.2.1. Replacing the apiKey

You need to replace the apikey in the file.

Because an Ohloh api has a limitation of 1000 requests per day, you may need to apply several apikeys according to your needs.
2.2.2. Change the xml page you want to retrieve.
As we mentioned, we retrieve data from the xml page, so you should change the xml page according to your requirements.

2.2.3. Change the data you want to retrieve
You should change the definition of the data to be retrieved according to your needs. For example, in the following figure, we retrieve the data of “id”, “name”, “created_at”..., you should change the data according to your requirements.
for (int j = 0; j < resultElement.getLength(); j++) {
    Element element = (Element)resultElement.item(j);
    NodeList projectElement = element.getElemenbyTagNames("account");
    for (int k = 0; k < projectElement.getLength(); k++) {
        Element element = (Element)projectElement.item(k);
        String id = element.getElemenbyTagNames("id").item(0).getTextContent();
        String name = element.getElemenbyTagNames("name").item(0).getTextContent();
        name = name.replaceAll("","");
        name = new String(name.getBytes("UTF-8"));

        String created_at = element.getElemenbyTagNames("created_at").item(0).getTextContent();
        created_at = created_at.replaceAll("","");
        created_at = created_at.replaceAll("","");

        String updated_at = element.getElemenbyTagNames("updated_at").item(0).getTextContent();
        String description = element.getElemenbyTagNames("description").item(0).getTextContent();
        updated_at = updated_at.replaceAll("","");
        updated_at = updated_at.replaceAll("","");
2.2.4. Print out the data you want.

2.3. Compile and execute your java file

Run cmd and make the place where you locate your java file as the current directory.

Use javac Ohlohaccounts.java to compile the java file,

and use java Ohlohaccounts>E:\tutorial\accounts.txt to execute the file and save the results to accounts.txt.
After the execution is done, you will get an txt file name “accounts.txt” storing the data you have retrieved.

3. **Load data into MySQL database**

   Please refer to the tutorial 1: MySQL Database and Relational Data Processing