

Alchemy API keyword Extraction - Training Material | WP3

Author: Bob Boelhouwer INL internal review: Jesse de Does, Katrien Depuydt

Introduction

AlchemyAPI (<http://www.alchemyapi.com/>) provides an online service for keyword extraction, but also for sentiment extraction, text categorization, etc. That means that all tools are available as web services. There is no need to install software locally. This can be considered an advantage. But a disadvantage is that it is not possible to tweak and tune software for special purposes. Moreover, there is only a fixed number of language modules available for the services. For keyword extraction these are: English, French, German, Italian, Portuguese, Russian Spanish and Swedish. It is not possible to use language models for other languages or for special text types.

AlchemyAPI is a commercial product. A license key is required to use it. Luckily, it is possible to receive a free license (<http://www.alchemyapi.com/api/register.html>), as long as there are less than 1000 transactions per day.

Requirements

As was mentioned in the previous section, no software needs to be installed, but the workflow will require an internet connection and a facility to send and retrieve information over http. An example of a simple client application is [CURL](#), which is available for most LINUX or UNIX distributions.

Usage

Since web services can be applied to an unlimited number of purposes, we will not try to describe all possible applications. In this document we will explain how a request is compiled and how that request can be submitted to the web service with a simple command line tool like [CURL](#). In the context of a workflow the API will likely be accessed via scripting tools like PHP or Perl. For most of these tools there are modules available that can send and retrieve information over http. AlchemyAPI also provides toolkits for embedding the service in software solutions (<http://www.alchemyapi.com/developers/sdks/>).

The API is called using a REST interface. The documentation of the interface for keyword extraction is quite extensive, and can be found here: <http://www.alchemyapi.com/api/keyword-extraction/>.

Basically, the interaction consists of a string that is sent by an http-client to a server of AlchemyAPI. That server then performs an action on the data and returns a file with the result to the http-client. The string that is sent to the service is in a standard URL format like 'http://access.alchemyapi.com/calls/text/TextGetRankedKeywords'. GET request parameters are used to further instruct the web service. These will be explained below. There are three methods available that differ in the way the data is presented to the service.

1. Web API [URLGetRankedKeywords](#) is used for processing publicly-accessible Internet web pages.



2. HTML API [HTMLGetRankedKeywords](#) is used for processing uploaded HTML content.
3. Text API [TextGetRankedKeywords](#) is used for processing uploaded text content.

Since text API seems the most useful method to handle results from digitization processes, we will focus on that. Text API, however, has a limitation to the size of text documents that can be uploaded of 150kB.

The API provides a number of parameters. Below we will give a description of the most relevant ones (liberally taken from the AlchemyAPI website):

- **apikey**: this is the private key you obtain from AlchemyAPI (required parameter)
- **text**: Text document content (must be uri-argument encoded) (required parameter)
- **url**: Text document URL (must be uri-argument encoded). Simply adds this url to the output data. Can be used to hold an id to the original document or for tracking (optional parameter).
- **maxRetrieve**: maximum number of keywords to extract (default: 50) (optional parameter)
- **keywordExtractMode**: keyword extraction mode (normal or strict). Possible values: *normal* keyword extraction mode (default) or *strict* keyword extraction mode (returns more "well-formed" keywords). refines results at the expense of returning fewer keywords (optional parameter).
- **Sentiment**: whether to enable keyword-level sentiment analysis. Possible values: 1 – enabled, 0 - disabled (default) (optional parameter)

These parameters can easily be added to the request sent to the web service by e.g. typing the following string in the url field of a webbrowser:

```
http://access.alchemyapi.com/calls/text/TextGetRankedKeywords?
apikey=xxxxxxx&text=hello%20world&maxRetrieve=strict
```

The value of the parameter *apikey* needs to be, of course, a genuine license key. The value of *text* has to be url-encoded (like the space in the url above). Most scripting tools and utilities provide functions to do that. In the example below we will show how CURL will handle the encoding of text.

The output consists default of a xml-document. It can contain the following fields:

- **status**: success / failure status indicating whether the request was processed.
- **language**: the detected language that the source text was written in.
- **url**: the value of the url parameter that was previously sent to the service.
- **relevance**: relevance score for a detected keyword. Possible values: (0.0 - 1.0) [1.0 = most relevant]
- **sentiment**: sentiment for the detected keyword (sent only if keyword-level sentiment analysis is enabled)
- **statusInfo**: failure status information (sent only if "status" == "ERROR").

Example of use

Below is an example on how to use the web service with the unix tool 'curl'. Also here, the parameter 'apikey' needs to hold a genuine key from AlchemyAPI. Curl sends a http request to the url specified in the first argument. The parameter and value pairs to that request can be specified in a number of '--data' arguments. The argument '--data-urlencode' can be used if values need to be encoded.

```
curl http://access.alchemyapi.com/calls/text/TextGetRankedKeywords \
```



```
--data-urlencode "text=US Vice-President Joe Biden has urged Japan and South
Korea to 'improve their relations and co-operation'. Mr Biden was in Seoul on the
third leg of an Asian tour dominated by tensions over China's newly-declared air
defence zone. China's zone covers disputed islands controlled by Japan and an area
claimed by Seoul. Mr Biden said on Thursday China's move had 'caused significant
apprehension in the region.' Mr Biden discussed the zone with South Korean leader
Park Geun-hye. The issue of North Korea and ways to restart long-stalled nuclear
disarmament talks were also high on the agenda." \
--data "apikey=xxxxxxxxxxxxxxxxxxxxxxxxxxxx" \
--data "maxRetrieve=5" \
--data-urlencode "url=http://www.inl.nl" \
--data "keywordExtractMode=strict"
```

The result of this query looks like this:

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <status>OK</status>
  <usage>By accessing AlchemyAPI or using information generated by AlchemyAPI,
you are agreeing to be bound by the AlchemyAPI Terms of Use:
http://www.alchemyapi.com/company/terms.html</usage>
  <url>http://www.inl.nl</url>
  <language>english</language>
  <keywords>
    <keyword>
      <text>Mr Biden</text>
      <relevance>0.95167</relevance>
    </keyword>
    <keyword>
      <text>Vice-President Joe Biden</text>
      <relevance>0.816955</relevance>
    </keyword>
    <keyword>
      <text>newly-declared air defence</text>
      <relevance>0.582731</relevance>
    </keyword>
    <keyword>
      <text>nuclear disarmament talks</text>
      <relevance>0.582006</relevance>
    </keyword>
    <keyword>
      <text>Korean leader Park</text>
      <relevance>0.535901</relevance>
    </keyword>
  </keywords>
</results>
```

