

Online Dictionary APIs - Options and Thoughts

Introduction

Originally from a conversation about dictionary service options with Justin in regards to the UIO+ work that we anticipate using for the SJRK platform as learner scaffolding, I'm pulling together some links from quick research I did.

This is not about the following scenarios:

- Browser plug-ins that allow dictionary lookups
- Browsers (like Chrome) that support dictionary lookups

This is about available online APIs for dictionary word definitions, and libraries that access them or make them easier to use.

Wiktionary

Wiktionary supports the [standard MediaWiki API](#) - this basically lets you return page markup in a somewhat parseable format.

- Here is a gist snippet example of doing this: <https://gist.github.com/nichtich/674522>
- And some projects to make working with Wiktionary content more machine-readable:
 - <https://github.com/dkpro/dkpro-jwktl>
 - <https://github.com/Suyash458/WiktionaryParser>
 - <https://github.com/Patatruc/word-definition> (Javascript)

Open Source Libraries / Dictionaries

- <https://wordnet.princeton.edu/>
 - An example of using it from Node: <http://progur.com/2016/12/how-to-use-wordnet-in-nodejs-applications.html>
 - Their list of related projects is worth perusing: <https://wordnet.princeton.edu/wordnet/related-projects/>
- <http://www.babelnet.org/>

Commercial Services

Many dictionary API services exist, including from the well-known dictionary publishers. These are examples, not comprehensive:

- <https://developer.oxforddictionaries.com/>
- <https://www.dictionaryapi.com/>
- <https://www.wordsapi.com/> (based on wordnet)
- <https://www.collinsdictionary.com/api/>

Non-English Dictionary Services

This needs filling out! Wiktionary has non-English versions available such as (https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%9999accueil), but their quality is not certain.

Thoughts for Implementation in UIO

- Define a standard format for dictionary entries - a JSON structure that converts to display HTML
- An integrator would need to wire up a preferred service returning that format
 - For the SJRK platform, we could create a Kettle-based dictionary datasource doing this as an example
 - It might be possible to implement an in-browser parser doing the same - not sure about how this might interact with same origin policies

Working with Natural Language in Javascript

Some natural language libraries, like <https://github.com/NaturalNode/natural>, may have dictionary features.