

## Scraping Japanese Kanji for an Educational Game

Dustin Seltz • Mengdi Wei • Jimmy Wang • Xudong Guo

### Questions:

- What are the most common characters, and information about them such as: frequency of use, meanings, readings, number of strokes, radicals, and difficulty level.
- A single kanji character could have various readings when formed into different words. Given a kanji, what words can it form, and what are the readings of this kanji in each of those words?
- (Visualization) WaniKani is a Japanese learning website with difficulty levels for each kanji character. Do the frequency of use, number of strokes, and grade levels have some relationships with the WaniKani levels?
- Taking into account the student's progress and goals, what is the best set of kanji / vocab to teach to them next?

### Using data from:

- [Jisho.org](http://jisho.org)  
A dictionary with all sorts of information for each kanji character.
- <http://genki.japantimes.co.jp/self/genki-kanji-list-linked-to-wwkanji>  
Additional information regarding difficulty level.
- [https://en.wikipedia.org/wiki/List\\_of\\_j%C5%8Dy%C5%8D\\_kanji](https://en.wikipedia.org/wiki/List_of_j%C5%8Dy%C5%8D_kanji)  
Information from the official Jōyō table.
- <https://scriptin.github.io/kanji-frequency/>  
A comparison of Kanji frequency from various sources.
- The app  
The user selects their goal to determine the order of kanji to learn.

### Code references:

- Another UCSC student, Bryan, started some code for Kanjirer:  
<https://colab.research.google.com/drive/1ulFNiml9YYQF2K5CXvri8PoeD20TicbK>

### Viz references:

- <https://scriptin.github.io/kanji-frequency/>  
A comparison of Kanji frequency from various sources.