Jsoup Demo

jsoup is a Java library for working with real-world HTML. It provides a very convenient API for extracting and manipulating data, using the best of DOM, CSS, and jquery-like methods.

jsoup implements the WHATWG HTML5 specification, and parses HTML to the same DOM as modern browsers do.

- scrape and parse HTML from a URL, file, or string
- find and extract data, using DOM traversal or CSS selectors
- manipulate the HTML elements, attributes, and text
- clean user-submitted content against a safe white-list, to prevent XSS attacks
- output tidy HTML
-

jsoup is designed to deal with all varieties of HTML found in the wild; from pristine and validating, to invalid tag-soup; jsoup will create a sensible parse tree.

Example

Fetch the Wikipedia homepage, parse it to a DOM, and select the headlines from the In the news section into a list of Elements:

```
Document doc = Jsoup.connect("http://en.wikipedia.org/").get();Elements newsHeadlines = doc.select("#mp-itn b a");
```

http://jsoup.org/download

http://jsoup.org/cookbook/extracting-data/attributes-text-html

http://jsoup.org/apidocs/

http://www.chrisdadswell.co.uk/android-coding-example-authenticating-clientlogin-google-reader-api/