

GoogleAPI

Google API

- Allows search over Google's entire index of billions of web pages
- Uses SOAP - Simple Object Access Protocol
 - used for information exchange (esp. http)
- But:
 - Limit of 1,000 queries per day (without special arrangement)
 - Limit of 10 results per query (top 10)
 - Can be overcome somewhat through query refinement)

Getting Started

- Register
- Agree to terms and conditions
- Get key
 - A 32 byte code
 - Needed for each query submission
 - E-mailed to you
- Download API

Google Java API

- API for Java is an executable jar
- Set appropriate classpath
- Some sample code:

```
import com.google.soap.search.*;
...
GoogleSearch googs = new GoogleSearch();
googs.setKey(myGoogleKey);
googs.setQueryString("TREC 2005 papers");
GoogleSearchResult googResults = googs.doSearch();
GoogleSearchResultElementr.getResultElements();
max = googResults.getEstimatedTotalResultsCount();

GoogleSearchResultElement myresults[] = googResults.getResultElements();
String a_URL = (String) myresults[1].getURL();
```

Some notes

- Limit is 1,000 queries
 - Catch GoogleSearchFault, or,
 - Local counter to track
 - In any case, put thread to sleep for 24 hrs. (or abend)
- Google server may get overloaded
 - In error processing, capture "Please try again in 30 seconds"
 - Put in 30 second sleep, and retry query

Some Notes

- `getEstimatedTotalResultsCount`
 - `max = googResults.getestimatedTotalresultsCount();`
 - max could be more than 10
 - max is an estimate; could be wrong
- Search only over main index (no GoogleScholar, Froogle, etc.)

API URL

- Site URL:
<http://www.google.com/apis/>
- AltaVista Search Developer's Kit
http://www.upenn.edu/computing/web/webteam/md/SDK/sdk_ref.htm
- Yahoo! Search Development Kit
<http://developer.yahoo.com/search/index.html>