

Project 2

Advanced Databases

Amandeep Singh
as3947@columbia.edu

Evangelia Sitaridi
es2996@cs.columbia.edu

March 17, 2011

1 Example

To test the classification results of QProber we used the provided the examples in the project description and several of the examples from a list of Largest Deep Web Sites [1].

Appendix

File Listing

We list below the files included in our submission zip. The zip contains the following folders:

[src, lib, queries]. The files are:

- README.pdf
- src/YahooProber.java
- src/Makefile
- lib/json.jar
- queries/Root.txt
- queries/Computers.txt
- queries/Health.txt
- queries/Sports.txt

Compilation

- **Compile:** make
- **Clean-up:** make clean
- **Execute:** java -cp ../lib/json.jar YahooProber *< url >* *< t_{es} >* *< t_{ep} >* *< yahoo - app - id >*

- **Yahoo App Id:**

BEWTNqTV34H1z0jJNQ5MZB48A1vR2mJeNAhKRvk5.bLyZd6gYgQmsVVsqZ7vv32aW73O6VNyzTO

| URL | Description | Classification |
|-----------------------------|-------------------------------------|----------------------------|
| java.sun.com | Java@Sun | Root/Computers/Programming |
| yahoo.com | Yahoo SE | Root |
| diabetes.org | American Diabetes Assoc. | Root/Health |
| tomshardware.com | Tom's Hardware | Root/Computers/Hardware |
| hardwarecentral.com | PC Hardware Reviews | Root/Computers |
| espn.com | WorldWide Leader in Sports | Root/Sports/Basketball |
| portal.acm.org | ACM Digital Library | Root/Computer/Programming |
| hopkins-aids.edu | HIV guide | Root/Health/Diseases |
| agiweb.org | American Geological Institute | Root/Health |
| www.cancer.gov | National Cancer Institute | Root/Health/Diseases |
| www.ncbi.nlm.nih.gov/PubMed | PubMed | Root/Health/Diseases |
| www.ovid.com/site/index.jsp | Ovid Technologies | Root/Health/Diseases |
| soccernet.espn.go.com | Football News & Scores | Root/Sports/Soccer |
| www.jumbo.com | Free Computer Software | Root/Computers/Programming |
| www.webmd.com | Medical News & Information | Root/Health |
| www.fitnessmagazine.com | Fitness Magazine | Root/Health/Fitness |
| www.afaa.com | Aerobic & Fitness Info of America | Root/Health/Fitness |
| processing.org | Processing Software | Root/Computer/Programming |
| www.telegraph.co.uk/sport | Telegraph Sport News | Root/Sports/Soccer |
| nba.com | National Basketball Association | Root/Sports/Basketball |
| www.sports.org.au | Australian Athletes with Disability | Root/Sports |

References

- [1] Largest Deep Web Sites, http://aip.completeplanet.com/aip-engines/help/largest_engines.jsp