Afshin Jamali

CSC 575

Project Proposal

**Build your own search engine**

The intent is to build a search engine with high precision with the use two or three domains, such as animal, car, medical, etc. Using polymorphism during indexing documents allows for terms to be indexed according to their specific domain.

The question becomes which domain the term is regarded as. The current solution I have is prior to indexing, determine the domain for each document manually and then feed it through the appropriate indexing criteria.

Indexing would be done through use of hashmap or dictionary. When a user searches for a term, and is looking for a term within a specific domain, they would type the word followed by the name of the domain. For example, “Jaguar animal.”

The search engine would contain the following elements:

* Query interface
* Read in documents and parse them using inverted file indexing format
* Relative term frequency information
* Utilize existing stemming and stop list tools
* Weighting schemes such as TF-IDF
* Vector-space model with comparing Cosine, Dice’s, and Jaccard’s similarity measures.
* Save the index to an offline storage such as database