CMPE297

**Design Document for Distributed Web Crawler and Indexer / P2P Map reduce**

Fall 2012

Chunyang Xia

Diego J Medina

Horng Lin

George Wang (georgeatgithub)

***Functional Requirements:***

1. As a user, I want to crawl a set of web sites so that I can do data mining on the web sites’ contents.
2. As a user, I want to index data from the crawler so that I can do keyword search on the contents.

***Description:***

The objective is to build distributed crawler that can crawl any sites and index into a search engine like Lucene.

**User story**

xxxxxxxxxxxxxxxxxxxxx

***Overall architecture Diagram***

(capture all key components of the system, and all interactions among components including inputs and outputs to each component)

***API Documentation***

i. Input parameters in the request

ii. Output parameters or payload structure in the response

iii. A list of error codes

iv. Request/Response example for each API