ReferenceStringCleaner

This class cleans the word that is passed in by DocParser. It checks if it is a stop word, removes punctuation, makes it lower case, and stems it to the root word. [Owner: Owen]

-AVLTree<string> stopwords;
-ifstream fin;
-char character;
-string now;
-string final;
-int num;
-unordered_map<string, string> map;
-string beforeStemming;
-bool test;
-AVLTree<StemmedWord> stemmedTree;
+ReferenceStringCleaner();
+void clean(string& needsCleaning);
+void removeStopWords();
+bool checkPunctuation(char& cha);
+void turnToStem(string& dirty);
+bool checkStopWord(string& dirty);
+stringstream word;

DocParser

This class parses the xml file by using rapidXML. it also relies on ReferenceStringCleaner to clean the words that it reads in. It stores all of the words that are read into the program and their frequencies on each page and adds them to the proper IndexInterface. [Owner: Owen]

-bool structureBuilt;
-ifstream fin;
-IndexHandler* indexHandler;
-string xmlFileName;
-stringstream stream;
+DocParser();
+DocParser(char* fileName, char* output);
+void setFilePath(string);
+bool isStructureBuilt() const;
+void clearStructure();
+void parse();
+void searchFor(string);

StemmedWord

This class is a container class that holds two strings. the first string is an unstemmed word and the second is the stemmed word. It is used to lookup a word in an AVL tree in ReferenceStringCleaner in order to make parsing faster. If a word is found in the tree, the second word is assigned to it because that is the result of stemming it. Looking up a word in an AVL tree is faster than stemming a word. [Owner: Jason]

+string beforeStemming;
+string afterStemming;
+StemmedWord();
+StemmedWord(string, string);
+bool operator<(const StemmedWord&) const;
+StemmedWord& operator+=(const StemmedWord& rhs);

RapidXML

We use RapidXML to parse xml files. It was downloaded from the website below.

http://rapidxml.sourceforge.net/

Porter2_Stemmer

We use Porter2_Stemmer to stem the words from the file. It was downloaded from the link below.

https://bitbucket.org/smassung/porter2_stemmer/src/078e3a0fea8713b68520332955dc5280a20bb162/porter2_stemmer.cpp?at=master