Web Crawler Design Grading Sheet

Name: _____

TA: _____

Score	Max Possible

Data Structures

Detailed description of the following data structures, including justification for each choice

_____	**D**ata structure for tracking unprocessed pages
_____	**D**ata structure for tracking already processed pages
_____	**D**ata structure for managing stop words
_____	**D**ata structure for storing word/page index

Class Responsibilities

_____	Drive the overall crawling process
_____	Store the URL and description for a page
_____	Store index that maps words to pages
_____	Keep track of yet-to-be indexed pages
_____	Keep track of already indexed pages
_____	Load and store stop words
_____	Distinguish between HTML and non-HTML links
_____	Distinguish between in-scope and out-of-scope links
_____	Resolve relative URLs
_____	Parses words, links, and description from HTML pages
_____	Populate word index
_____	Generate XML output file

Algorithms

Top-level code for the following algorithms:

_____ Main driver for the crawling process

_____ HTML parser

_____ XML output page generation

Design Quality

_____ Cohesive classes and methods

_____ Effective information hiding

_____ Effective class, method, and variable names

_____ Clear, easy-to-read document

_____ **Total**