Ain Shams University
Faculty of Computer and Information Sciences
Data Strucutres   Project

Title	Search Engine Tokenizer
Description	-It's a program that takes an HTML file and divide it into tokens (sentences and words).
Group size	2 – 3 members.
Duration	3 weeks.
Deliverables	1- Deliver a list of words which are used in each input HTML file.
Bonus extensions	-Add the location of each word in the output list.
Mentor	TA Sherif Haggag
Notes	Check http://en.wikipedia.org/wiki/Web_search_engine for more info