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

Title	Web Crawler
Description	• A **Web Crawler is a software for downloading pages** from the Web • Also known as Web Spider, Web Robot, or simply Bot Cycle of a **Web crawling process:** • The crawler start downloading a set of **seed pages, that are** parsed and scanned for new links • The links to pages that have not yet been downloaded are added to a central queue for download later • Next, the crawler selects a new page for download and the process is repeated until a stop criterion is met
Group size	3 – 4 members.
Duration	4 weeks.
Deliverables	1- A Web Crawler which takes the starting URLs and crawls the internet till it saves 1000 different HTML pages.
Bonus extensions	-Save the HTML page without HTML tags.
Mentor	TA Sherif Haggag
Notes	Check http://en.wikipedia.org/wiki/Web_search_engine for more info