School of Electronics And Computer Science
ELEC6050   MEng Group Design Project
_____
Project Specification And Plan

Title: Trending Terms Malware Crawler

Supervisor: Tim Chown

Team Members: Thomas Grainger

Weike Liao

Chris Orchard

Nafiseh Vahabi

Customer: Team Cymru (http://www.team-cymru.org)

Project Specification:

The system should ascertain trending search keywords by searching Google Trends, Twitter's trending topics for a specific WOEID (Where On Earth ID), RSS feeds such as tabloids.

Each of these keywords found can then be used to determine a set of relevant website URLs. To do this a set of search engines, such as the search engine aggregator http://dogpile.co.uk, are queried with each of the keywords found.

The URLs discovered should then be dispatched to a distributed set of workers using a Task Queue. Each of these workers then download and examine the documents referred to by each of the URLs.

Care must be made to ensure that these robot workers look like humans browsing the web: techniques such as using a wide range of IP addresses, emulating the Referer and UserAgent request headers and even running instances of actual browsers, such as Firefox and Chrome. Research should be made into how existing malware distribution networks evade detection, so as to counter those systems.

The results of these workers should be stored in a database so as to be later retrieved by a web application reporting system to be displayed to the end user.

Name	Work	Oct 2012		Nov 2012									Dec 2012			Jan 2013			
		Week 40	Week 41	Week 42	Week 43	Week 44	Week 45	Week 46	Week 47	Week 48	Week 49	Week 50	Week 51	Week 52	Week 1	Week 2	Week 3	...	
Project Startup	**11d 1h**																		
Prepare Project Brief	**8d**																		
Write Brief	4d																		
Build Gantt Chart	4d																		
Agree Brief With Customer	3d																		
Submit Brief	1h																		
Development Iteration 1	**49d**																		
Retrieve Key Words	7d																		
Get Search Results	7d																		
Render Web Page Content	14d																		
Detect Malware	7d																		
Store Detection Results	7d																		
Produce Detection Report	7d																		
Progress Seminar 1	**5d**																		
Prepare Presentation	3d																		
Practice Presentation	2d																		
Seminar																			
Development Iteration 2	**28d**																		
Render Web Page Content	7d																		
Detect Malware	14d																		
Produce Detection Report	7d																		
Development Done																			
Progress Seminar 2	**5d**																		
Prepare Presentation	3d																		
Practice Presentation	2d																		
Seminar																			
Project Report	**21d**																		
Write Report	21d																		
Project Report Handin																			
Final Presentation	**6d**																		
Prepare Presentation	4d																		
Practice Presentation	2d																		
Presentation																			