

Intern Requirement

No. of students required	1 to 2
Preferred Discipline	Information Technology
Prerequisites/ Skills Required	Familiarity in: <ul style="list-style-type: none">• data-mining from the web• text-processing, e.g. Regular expression and XSLT

Project Details

Title	Data Extraction from Internet
Overview/Background	With the arrival of Web 2.0, it is much easier to publish to the web now than before. This has in turn motivated more journalists and columnists to publish their analysis online rather than through the conventional paths.
Objectives/Scope/ Deliverables	<ul style="list-style-type: none">• Implement an open-source based solution to mass extract data from the web. Subsequently, to re-organise the retrieved data - by sieving out the necessary ones - into meaningful information.• Upon completion of this project, the intern will acquire good knowledge and hands-on experience in enterprise Java application and web-mining techniques
Project Duration	6 months