

**Intern Requirement**

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

**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 motivated more journalists and columnists to publish their analysis online rather than through the conventional paths.
Scope	This project involves implementing an open-source based solution to grab the webs for quality information on a massive scale. After the grabbing process, additional processing is required to sieve essential data from the bulk unnecessary information gathered during the process.
Aims/Objectives/Deliverables	<ul style="list-style-type: none"><li>• To implement an open-source based solution to extract quality information from internet sources and re-organize them into meaningful entity.</li><li>• Upon completion of this project, the intern will acquire good knowledge and hands-on experience in enterprise Java application and web-mining techniques.</li></ul>
Project Duration	6 months