

Education

Masters in Technology in Information Technology, 2004-2006.

Kanwal Rekhi School of Information Technology,
Indian Institute Of Technology Bombay, Mumbai, India.
Current CPI - 9.12 (semesters 1-3).
Current CPI - 9.3 (semesters 1-3, and MTP first stage).

Bachelor of Engineering in Computer Science and Engineering, 2000-2004.

Shri Guru Gobind Singhji College of Engineering and Technology, Nanded, India.
Aggregate Percentage - 81.6%

12th, 2000.

Kola Perumal Chetty Vaishnav Senior Secondary School, Chennai, India.
Aggregate Percentage - 89.2%

10th, 1998.

Kola Saraswati Vaishnav Senior Secondary School, Chennai, India.
Aggregate Percentage - 81.2%

Projects, Seminar, & Termpapers

M.Tech. Project

Web based Named Entity Recognition

Supervisor : Prof. Sunita Sarawagi, KReSIT

Web-NER aims at extracting entities of interest from web pages. The scale, unstructuredness, and diversity of the web pose challenges to NER on the web-pages. Traditionally, rule based techniques like Wrapper Induction Systems have been used for this task but these techniques are site specific and not robust. We intend to use statistical learning based approaches.

Our task is learn to a model that could label blocks on a web-page. The blocks have rich formatting and layout properties, in addition to textual properties, and the model must use these cues for labeling blocks as (sub)entities. There are spatial relationships between (sub)entities and these can provide vital clue to the model. The blocks, however, can be of varying sizes and are irregularly placed on the 2D layout of a web-page, and this makes capturing contextual interactions (spatial dependencies) between blocks is a challenging task.

In this project, our aim is to build a framework that will assist in entity extraction from web-pages by exploiting textual, visual and spatial properties. We concentrate mostly on entities composed of several sub-entities that are dispersed on a web-page. In our initial attempts, we have used CRFs and SVMs with simple textual, spatial and visual features. For our experiments, we found that SVMs perform better than CRFs.

May 2005 – Present

KReSIT, IIT Bombay, Mumbai

M.Tech. Seminar

Application of Instance based Learning for

Damage Detection in Mechanical Structures

Supervisor : Prof. Sunita Sarawagi, KReSIT

Here, we surveyed some of the classification and clustering algorithms and discussed an application involving use of Instance-based learning techniques for damage detection in mechanical structures where huge amount of data is available and where the number of target variables to be predicted are also very high.

July 2004 – November 2004

KReSIT, IIT Bombay, Mumbai

B.E. Project

Design and Implementation of a Native XML

Filesystem

Supervisor : Mr. P.S Nalawade, SGGSCoET

We designed and gave a proof of concept implementation of a Native XML Filesystem. Such a filesystem is XML aware in sense that it is specially designed for natively storing XML and provides logical access to XML data. Native XML databases ignore the underlying storage model and care for providing only logical access, and hence, performance of applications using these databases suffer. We feel that a solution for storage and access of XML forms a building block of many big XML systems, and hence it must be at filesystem level. In our approach, the storage and access models are inspired by *XISS* system.

January 2004 – June 2004

SGGSCoET, Nanded

- Implementation Techniques in DBMS - Implementation of left-deep tree based Query Optimization for light weight DBMS(Delite) for handheld devices using C.
- Object Oriented Technology - Design and implementation of Query Engine of XMLSpaces - An Associative Distributed Shared Memory System for XML documents. We defined our own Query language which was similar in some aspects to XPath. Tools used - Java, JavaCC, JUnit, XML, XPath, XQuery.
- Course termpaper in Object Oriented Technology - Applications of Aspect Oriented Programming to Middleware Systems.
- Principles of Distributed Computing - Dependency Based Root Cause Localization.

January 2005 – April 2005

Second Semester

- Systems Lab - Secure Grade System using the Smart Cards in circulation at IIT Bombay. It allows a professor to sign the grades which are then stored in an untrusted database. These grades can later be verified for checking any tampering. Tools used - Java, Java WebStart, PC/SC Smart Card API and its JNI wrapper.
- Data Mining - Implementation of the Scalable EM clustering algorithm in Weka, a popular Data Mining platform. Tools used - Java, Weka.
- Web Mining & Information Retrieval - Study and implementation of Blelloch's document-id reordering algorithm for improving gamma-coding based inverted index compression. Tools used - Lucene.
- Course termpaper in Principles of Distributed Computing - Dependency Based Root Cause Localization.

Skills

- *Operating Systems:* Linux, Windows
- *Programming Languages:* C, C++, Java
- *Scripting Languages:* Perl, PHP, Shell Scripting
- *Databases:* MySQL, PostgreSQL
- *W3C Technologies:* HTML, DHTML, CSS, Javascript, XML, XSchema, XPath, XSLT
- *Tools:* Matlab, Linux Tools

Summer Internship

At *Deusco Technologies, IIT-Bombay*, in June, 2002 - July, 2003. We were involved in building of a Web based Widget Toolkit.

Academic Achievements

- GATE-2004 percentile was 98.83 with an All India Rank of 406.
- Qualified NCST G-Level examination-2004 and was among the top 2 of the country.
- Qualified NCST D-Level examination-2001 and secured 95 percentile.
- Topped University in 2nd and 4th Year of B.E.(CSE).

Extra Curricular Activities

- Was the Cultural Secretary in year 2002-2003 at S.G.G.S. College of Engineering and Technology.
- Was the one of the editors for QUASAR - the technical journal of dept. of C.S.E at S.G.G.S. College of Engineering and Technology, Nanded.
- Was the web admin in year 2003-2004 at S.G.G.S. College of Engineering and Technology.
- Won prize at College level software contest in 2003-2004.
- Selected for 10 day NSS camp in 2002-2003. Organized few blood donation camps under NSS.

Areas of Interest Datamining, Information Extraction, Information Retrieval, Databases, Algorithms, and XML Technologies.

Personnel Information Name : Sandesh Tawari
DOB : February 04, 1983
Gender : Male
Father Name : Mr. Damodardasji Tawari
Address : Lab SIC312, KReSIT, IIT Bombay, Powai, Mumbai - 400076, India.
Telephone : +91-22-2576-4974, +91-9869176822
Email : sandesh@it.iitb.ac.in
Home page : <http://www.it.iitb.ac.in/~sandesh/>

Declaration The above furnished information is true to the best of my knowledge.

Sandesh Tawari
December, 2005
IIT Bombay, Mumbai.