

EDUCATION

Master's in Computer Science

Expected Fall 2011

GPA 4.0, Stony Brook University

Courses Operating Systems, Analysis of Algorithms, Concurrent and Distributed Algorithms, Natural Language Processing, Computational Biology, Words and Pictures.

Master's Project Text mining for big-data analysis, under the advisement of Prof. Steven Skiena

Bachelor of Engineering (Computers)

Jun 2006

First Class, Vidyavardhini College of Engineering and Technology

Courses Artificial Intelligence, Advanced Databases, Operating Systems, Computer Networks

SKILLS

Languages

- Proficient in Java, Python, C#, Javascript, Bash, C++

Data

- Classification and mining with Weka, numpy, NLTK, learning R
- Data mining on Microsoft Analysis Services with MDX/DMX, SQL

Cloud

- Google AppEngine applications, using webpy and Django
- Amazon EC2 for distribution (via Hadoop)

Frameworks

- OR Mapping frameworks (NHibernate, Microsoft Entity Framework, Gentle.Net)
- Social frameworks (GData Python Client, Facebook applications)
- Web frameworks (ASP.Net, Django, Google AJAX API, jQuery)

CAREER

Software Developer Intern, Kindle Content Quality Team, Amazon Inc.

May 2011 - August 2011

Improved the Kindle experience by understanding user behavior using machine learning

Technical Leader, Neorithm Technologies

Oct 2008 - July 2010

Core team member since inception, led the technology R&D across the company's product lines

Senior Software Engineer, Finacus Solutions

Nov 2007 - Aug 2008

Core team member, enabled rapid application development with an in-house framework

Subject Matter Expert, Amdocs Development Center (India)

Aug 2006 - Oct 2007

Implemented Amdocs' telecom billing functionality using Unix processes

PAPERS AND PUBLICATIONS

- Language of Vandalism: Improving Wikipedia Vandalism Detection via Stylometric Analysis — **ACL 2011**
- Domain Independent Authorship Attribution without Domain Adaptation (**presenter only**) — **RANLP 2011**
- Filed for a patent on techniques used at Amazon during internship

PROJECTS

Vandalism detection in Wikipedia

Stony Brook University

Hypothesised and validated an NLP approach to detecting vandalism, by modeling good and vandal edits using a probabilistic context free grammar, and evaluating new edits, attaining 84% precision and 89% recall.

Pattern recognition in human speech

Stony Brook University

Built a classifier to detect the language of a human speaker, by building an ngram model over various features of an audio signal. The classifier attained an accuracy of 72% and recall of 73%, when tested over 9 languages.

Language detector for textual data

Stony Brook University

Implemented a novel algorithm, under the guidance of Yuri Puzis, to **detect the language of a given piece of text using a trie based decision tree**. The accuracy of the classification could be increased to an arbitrary level – limited by a dictionary – while training the decision tree.

Social network based on recommendation**Self and friends**

Built a **recommendation based social network on the Google AppEngine**, using Django. Given things one likes and doesn't like, people whose likes were most similar were matched up.

Differential AJAX Server**Neorithm Technologies**

Modified the ASP.Net pipeline to AJAX-ify legacy ASP.net applications, based on the 'diff' of the pages being transmitted to the client. Built custom modules such that the server would only hand out a diff of the new page with respect to the older page, and an injected javascript on the client would merge the changes.

Optimized MDX query generator**Neorithm Technologies**

Developed a **translator to generate an optimized MDX query** for a given high level declarative user query in XML, which was used to query a Microsoft Analysis Services OLAP warehouse. This was part of an Anti-Money Laundering product, whose development I led.

Versioned file storage engine**Neorithm Technologies**

Developed a **datastore to store files**, and exposed an API in C#. The datastore had versioning support, and supported multiple implementations for actual storage of files. In particular, a file-system based, and a relational database backends were built.

ATM switch**Finacus Technologies**

Built a **daemon to run on ATMs, and carry out transactions with banks** using the Diebold and NCR protocols. The daemon was written using the Microsoft.Net framework and was extensible to support other protocols.

Web application development framework**Finacus Technologies**

Built a **web application development framework**, consisting of UI elements in the form of ASP.net controls, an ORM, workflows and code generators. These greatly reduced the development time and provided independence from 3rd party vendors.

Rater extensions**Amdocs**

Developed various daemons and C++ extensions to support the telecom billing processes of Amdocs' *Rater* product, on Sun Sparc and AIX systems. Also traveled to Israel and Russia for testing and production support.

The Distributed DataBase**Vidyavardhini College of Engineering**

Developed the parser, query execution engine and GUI for **The Distributed DataBase** (<http://tddb.sf.net>), **a true distributed database built from the ground up**, which conforms to the SQL-92 syntax.

Email Server**Vidyavardhini College of Engineering**

Built an email server in JSP/servlets, where the communication between the server and client was done through web services.

Wordlist trainer**Self**

Built a web application <http://wl.sandesh247.com> to help students prepare for wordlists for exams like the GRE. It uses an **AJAX interface which communicates with the backend using a REST-api**, and has unique approach to learning - *revise that which you did not know the last time*, taking out the tedium to manually mark "hard" words.

Various jQuery plugins**Self**

Built various **jQuery plugins** during the course of my employment in Neorithm Technologies, while improving usability in web applications. Some of them can be seen in online repositories linked at the end of this resume.

LINKS

- Code samples: <http://github.com/sandesh247>; <http://code.google.com/u/sandesh247>
- Blog: <http://sandesh247.com/journal>