

VISHNU VYAS SETHUMADHAVAN

287 Laurie Meadows Drive, Apt 131
San Mateo, CA 94403

Tel : (213)-321 0411
Email : me@vishnuvyas.com
<http://vishnuvyas.com>

OBJECTIVE

To become an accomplished data mining and NLP engineer.

EDUCATION

2007 – 2008 University of Southern California Los Angeles, CA
■ M.S in Computer Science.
■ Courses Taken : Database Systems, Empirical methods in NLP, Convex Optimization and Theory of Computation (GPA : 3.5)

2002-2006 SRM Engg. College, Anna University Chennai, India
■ B.E in Computer Science.
■ Graduated with distinction (82%)

INTERESTS

- Large Scale Information and Knowledge Extraction.
- Statistical Methods and Machine Learning.
- Computational Semantics (particularly, Lexical Semantics)
- Modeling and Explaining Semantic Similarity

WORK EXPERIENCE

March 2008 – July 2009 Yahoo! Inc. Santa Clara, CA
Research Intern

- Implemented a system for summarizing attributes and instances from a concept automatically extracted from web text (ala Google Squares).
- Implemented a system for helping human editors create better seed sets by automatically removing less impactful seeds.
- Developed two algorithms for refining sets of entities using minimal annotation effort. This work led to a paper published in NAACL-09.
- Implemented the two algorithms on the Yahoo Grid using Hadoop and pymapred.

July 2007- March 2008 Information Sciences Institute Los Angeles, CA
Research Assistant

- Formulated and Implemented an Information theoretic measure for managing diversity in similarity explanations.
- Implementation of a system to automatically cluster people using the email body text and sending and receiving patterns.

December 2005 – February 2006 CAIR, DRDO Bangalore, India

Developer

- Part of a team involved in the design, development and testing of Swarm Algorithms for site selection in GIS (This work led to a paper on the same topic published in CEC 2006).
- Deployment of a Web Application (using Tomcat) for combining multiple sources of information along with GIS information.

March 2005–December 2005 CAIR, DRDO Bangalore, India

Part time / Student Intern

- Design and development of a framework for swarm algorithms in python.
- Implementation of a Domain Specific compiler for converting fixed format text to XML (SVG).

PUBLICATIONS

- Patrick Pantel, Eric Crestan, Arkady Borkovsky, Ana-Maria Popescu and Vishnu Vyas, Web-Scale Distributional Similarity applied to Entity Set Expansion. - EMNLP 2009, Singapore
- *Semi Automatic Entity Set Refinement* – Vishnu Vyas, Patrick Pantel – NAACL 09, Boulder, CO.
- *Explaining Similarity of Terms* – Vishnu Vyas, Patrick Pantel – COLING 2008, Manchester, U.K.
- *A Joint Information Model for n-best Ranking* – Patrick Pantel, Vishnu Vyas – COLING 2008, Manchester, U.K.
- *An Algorithm for Site Selection in GIS based on Swarm Intelligence* – Ajay Sharma, Vishnu Vyas, Dipti Deodhare – IEEE Congress on Evolutionary Computation 2006.
- *Compression of Grayscale Images using Image Specific encoding based on Unsupervised Hebbian learning* – Bharadwaj Srinivasan, Vishnu Vyas. Published as a student paper in the Proceedings of the International Conference on Human Machine Interfaces, 2004.

TECHNICAL SKILLS

- Proficient in C/C++, Java and Python.
- Experience with distributed platforms such as Hadoop and MPI
- Proficient in implementing and using various machine learning algorithms such as HMM, Linear chain CRF and MaxEnt.
- Knowledge of Lexical Semantics, including distributional similarity, dynamic similarity modeling and set expansion.
- Experience with developing large scale (web-scale) algorithms for information extraction.

OTHER SOFTWARE PROJECTS

Forward Backward Algorithm for EM (library).

Description : An efficient C++ based implementation of the forward backward algorithm for hidden markov models. Can also be extended in use for Conditional Random Fields when using only local features.

Features :

- Uses log-space calculations to avoid overflow and uses linear addressing for 2D arrays thus increasing efficiency.
- Is extremely fast.

SBD (library)

Description: A C++ library for efficient heuristic sentence boundary detection and unsupervised phrase chunking.

Features:

- Hueristics tuned for English sentence boundary detection.
- Unsupervised phrase chunking detects X-bar style specifiers and heads of phrases using bigrams and a poisson prior.

Simple Finite Mixture Models (library)

Description : A simple python based implementation of estimating the parameters of a finite mixture model using both EM and Gibbs sampling.

Features :

- Uses SCIPY to efficiently sample from multivariate Gaussians.

The Cactus Virtual Machine

Description : A virtual machine for high level languages with support for dynamic typing and arbitrary precision arithmetic (using GMP). Also includes a compiler and a byte-code based static linker for a subset of the scheme programming language.

Website : <http://cactusvm.sourceforge.net>

Responsibilities

- Design and Implementation of a virtual machine in C with support for dynamic typing.
- Design and Implementation of a dynamic loader with preconfigured libraries in C. Design and Implementation of a static linker in python for resolving symbols during compile time.

HONOURS AND AWARDS

- Awarded the Indian President's Award for Best Student Project (2006)