KAPIL DALWANI

Email: kapildalwani@gmail.com **Web Page**: www.cs.jhu.edu/~kapild
3332 22nd Street, Apt B, S.F., CA: 94110

EDUCATION

- **Johns Hopkins University** — Baltimore, MD
  (M.S) Master of Science in Engineering (Computer Science) GPA – 3.54 — **1/08** to **08/09**
- **Panjab University, Punjab Engineering College** — Chandigarh, India
  (B.E.) Bachelor in Electrical Engineering, First class with Honors - 71.71%. — **2004**

CURRENT STATUS:

- Working as a Senior Software Engineer with Data Insight group at YellowPages.com. — 02/11 – Till date
- Software engineer for Search at Pipio Inc, San Francisco. (http://pip.io) — 12/09 – 10/10
- Graduated from Johns Hopkins University in MSE program of Computer Science. — 1/08 – 08/09
- 2 years experience working on Solr/Lucene and Hadoop/Hive stack.

MS COURSES:

Information Retrieval, Natural Language Processing, Data Mining, Machine Learning, Statistical Learning with Applications, Information Extraction from Speech and Text, Algorithms, Randomized Algorithm, Computer Vision.

PUBLICATIONS:

- ***N-gram Search Engine with Patterns Combining Token, POS, Chunk and NE Information*****,** Proceedings of LREC, 2010 by Satoshi Sekine, Kapil Dalwani
- ***New Tools for Web-Scale N-grams.*** Dekang Lin, Ken Church, Heng Ji, Satoshi Sekine, David Yarowsky, Shane Bergsma, z

Fun Projects:

a) Implemented a 3 legged OAUTH google chrome extension using Rdio API. http://bit.ly/mh9Q88
b) ToBikeToBart android App: https://github.com/kapild/ToBikeToBart

PROFESSIONAL EXPERIENCE

- ***YellowPages.com (AT&T Interactive), San Francisco*** — Senior *Software Engineer for Search/Data Insight 02/11 to till date*
  - Currenly, working with the Data Insight group to mine user query logs.
  - Implemented Related Search, Spell correction and Query Synonyms using Hadoop piplelines.
  - Implemented a data library to read data from hive tables and use it in map-reduce jobs for further processing.
  - Currently, buillding an offline simmulator for search by mining user query logs, helping predict the usuabilty of a new launch of search version hence minimizing the manual efforts required for A/B testing.
  - Worked with the Search relevancy team to improve relevancy and ranking of search results using SOLR and implement new features likes CamelCase, hours of operation of business, customizing lucene/Solr code.

- ***Pipio(http://pip.io), San Francisco*** — *Software Engineer for Search 12/09 to 10/10*
  - Worked on Lucene, SOLR and spearheading the index and search related development work.
  - Implemented cool stuff like Hashtags, Geohash related searches on the website.
  - Designed, analyzed, implemented and tested various functionality of product using PHP.
  - Handling the email servers, deployment of projects to live site using Python and other ad-hoc tasks.
  - Working on back-end services API, MySQL, XMPP, Pubsubhub and not so cool Java Script.

- ***CoreObjects, India*** — *Product Engineer 12/06 to 12/07*
  - **Siperian UI Toolkit:** Eclipse plug-in generating on the fly code, having features like yahoo-maps integration in Flex.
    - Designed, analyzed, implemented and tested code. Worked in a full Agile development life cycle.
  - **Facebar:** Audio file sharing site and live music streaming built on a thin social networking site.
    - Designed, analyzed, implemented and tested code. Implemented MVC architecture using Struts, Spring and Hibernate Search in a full development life cycle.

- ***Computer Sciences Corporation, India*** — *Software Engineer, 8/04 to 11/06*
  - **Online Examination**: In-house project making an online examination system.
    - Worked on Struts, Java/J2EE technology, JDBC. Designed the whole database in MySql and implemented cod

GRADUATE PROJECTS 1/08 to 8/09

- ***N-gram Search Engine**:*
  Developed a N-gram(7-gram) search engine on Wikipedia text while working at the CLSP workshop Summer '09. (http://www.clsp.jhu.edu/workshops/ws09) . I was a part of the n-gram team (http://www.clsp.jhu.edu/workshops/ws09/groups/ualkn).
  A running model can be found at here http://bit.ly/1pT4S0 and http://bit.ly/X6eKW. Final presentation: http://bit.ly/1Vx4UU

- ***Machine Learning Library:***
  Designed and implemented a library in Java for the following machine learning algorithms: Decision trees, Linear regression, Online Learning of Linear Classifier (Perceptron & Winnow), Non Linear Classification(Dual Perceptron with SVM kernels), Clustering(k-means algorithm).

- ***Decision-tree language model for English letters:***
  Implemented a decision-tree classifier for a 4-gram language model using cross validation.

- ***Early Parser on Treebank:***
  Implemented a priority agenda based chart parser for probabilistic context free grammar on Wallstreet corpus, parsing 10 sentences in just 90 sec

- ***Part of Speech Tagger:***
  Implemented Viterbi and Forward-backward algorithm to improve efficiency for POS tagging.

- ***Isolated word Recognizer:***
  Build an isolated word recognizer in speech using composite HMM and fenonic base forms.

- ***Text Categorization & Machine Learning on Newsgroups-20 dataset:***
  Used PCA and feature extraction analysis for text categorization on Newsgroups-20 & Movie dataset, subsequently analyzed the efficiency of various machine-learning algorithms using WEKA.

- ***Spam Detection using Back-off & Smoothing:***
  Implemented various n-grams, smoothing estimates and various back-off models like Good-Turing, Katz back off, Witten Bell for text categorization/spam detection/language modeling.

- ***Vector Model:***
  Implemented a vector-based IR engine similar to SMART system. Used the concepts of tokenization, stemming, stop-words etc. to calculate ranking of various documents and recall-precision at various levels.

- ***Lexical Ambiguity:***
  Implemented a vector-based classifier for resolving word sense and person-place ambiguity.

- ***Collection Fusion:***
  Implemented a Collection fusion based approach to gather search results from multiple search engines and in-turn rank the documents. Used WorldNet for query expansion and Robots for parsing the web.

TECHNICAL EXPERTISE

- Languages: C, STL, C++, Java, Perl, Python, PHP
- Web: J2EE, Solr, Lucene, Hadoop, Hive, SVMLight
- Other: MySql, MATLAB, Weka, SVN Tortoise, Mercurial, Git

AWARDS & ACHIEVEMENTS

- Ranked 2333/150000 candidates in All India Entrance Exam Indian Institute of technology – **IIT 2001.**
- Punjab Engineering College 2000: Received Full one year Scholarship for being amongst top 3% in class.

PERSONAL

- Single, willing to relocate and travel.
- Interest includes Biking, Cooking, Rock Concerts and playing Tennis.