



Alex Ksikes

alex.ksikes@gmail.com

homepage: alex.ksikes.net

CAREER GOAL My goal is to work with passionate like-minded people and to build great products that would be used by millions of people.

INTEREST Text Mining, Information Retrieval, Kernel Methods, Web Development, User Interfaces

EDUCATION [University of Cambridge](#), Department of Engineering, Oct 2005 - Now

Ph.D. in Computer Science, Machine Learning.

My advisor is [Zoubin Ghahramani](#)

Master of Engineering in [Computer Science](#), [Cornell University](#), June 2003

GPA = 3.92 / 4.0

B.Sc. in [Computer Science](#) with Minor in [Applied Math](#), [Cornell University](#), June 2002

Dean's list Spring 1999 and Fall 2002

Math Sup/Spé, Lycée Polyvalent Ort, Strasbourg, France, Sept 1997 - Jan 1999

Scientific Baccalaureate with a Mention, Lycée International, Strasbourg, France, July 1997

Graduated with highest Honors, Mayfair High School, Los Angeles, June 1995

Exchange student in an American high school, scholar athlete award.

EXPERIENCE **June 2009 - Sept 2009**

Microsoft Research, Redmond

Intern at [Microsoft Bing](#) working on **ranking** of web search results. **Ensemble methods** were performed on large library of models and on datasets of **millions of documents**. Our method showed a **significant improvement** of NDCG and is currently being tested for production use.

May 2006 - Sept 2006

University of California, Berkeley

Visiting Researcher, worked on **text mining** and search interfaces.

BioText Search Engine: beyond abstract search published in Bioinformatics.

June 2005 - Oct 2005

University of California, Irvine

Institute for Genomics and Bioinformatics

Visiting Researcher, worked on **kernel methods** for the analysis of **chemical information**.

One- to Four- Dimensional Kernels for Small Molecules and Predictive Regression of Physical, Chemical, and Biological Properties published in JCI.

June 2004 - Sept 2004

Intern, Xerox Research Center Europe, Grenoble

Here is a [presentation of Ensemble Selection](#) given at Xerox.

Jan 2003 - June 2003

- My master's final project: *Ensemble Selection from Libraries of Models* is published in ICML 2004.

- Teacher Assistant on [CS 478: Machine Learning](#) at Cornell.

- BOOM 2003 on [Extreme Ensemble Selection](#)

Sept 2002 - Dec 2002

Teacher Assistant on [CS 100: Intro to Computer Programming at Cornell](#)

June 1999 - June 2002

Math tutors for students taking the scientific baccalaureate.

SKILLS

Mathematics: Multivariable Calculus, Analysis, Linear Algebra, Abstract Algebra, Algorithmic.

Programming: Preferred programming language is Python. I also like to program in C, Java and Scheme. Fluent in Unix tools. Other skills in HTML, CSS, PHP and SQL.

Language: French (native) and English (fluent)

PROJECTS
research related

[Cloud Mining](#)

Cloud Mining automagically builds cool faceted search interfaces for your data. Simply get your data in a specific format, provide a custom look and feel for each search result and Cloud Mining does the rest.

The interface allows a user to dig through the results by adding more tags to his query. He can also search sideways for similar items by clicking on "more of this" (IMDb instance). Different visualizations (such as tag clouds or bar charts) could be chosen which provide insights of the results being searched and encourage exploration.

See it in action on [IMDb](#) (1M movies) or on [DBLP](#) (1.2M computer science references) or on the whole of [MEDLINE](#) (20M articles with abstracts).

[Biomed Search](#)

Biomed Search is the largest search engine to **look up images** specifically in the **biomedical domain**. The system searches within captions and referring texts to images. Currently over **one millions images** are indexed. The projects was created in 3 months of works. I used the Lucene retrieval engine and the web.py framework. I did everything on the site: setting up the server, parsing, indexing and web programming. All the back end work was programmed in python.

Here is a [presentation of Biomed Search](#) given at the University of Cambridge.

[Biotext Search Engine](#)

I also contributed to UC Berkeley's Biotext Search Engine. Search within captions and abstracts of biomedical images. More information on this [publication](#).

PROJECTS
web related

[MLSS Admin](#) - [\[source code\]](#)

The admin we used at Cambridge for the [summer school](#). It let us browse through the fast amount of candidates, comment on each applicant and assign scores. [Try a demo](#), [how to use it](#), [get the source code](#) on github. Feel free to use for subsequent summer schools.

Also available is a simple **mass mailer** ([try a demo](#), [source code](#)) used to email all the admitted students.

The admin has been succesfully used to organize the Machine Learning Summer School 2009 and 2010 so far.

[Chiefmall](#)

With **Chiefmall** you can search for a **contractor near you**. Over 420 000 contractors including **all active licensed contractors in California** are currently listed. Contractors who join in can easily build up their **own website** with a unique url, picture portfolio, multiple work locations and more. Users can contact these contractors or **post jobs** to them.

Here is an **example with explanation** of a [custom website](#). And here are some screenshots of a contractor's account: [dash board with todo list](#), [describe your company](#), [service locations](#), [the portfolio](#), [adding a caption](#), [get jobs](#) or [get your license verified](#).

It took me 6 months of work to create this project. I did everything on the site: setting up the server, crawling parsing the data, geocoding, UI, MVC design, web design, concepts and ideas and SEO. All the back end was programmed in python.

The website also **features an admin** created using a program that makes on the fly Gmail like looking interface. Some screenshots: [the state categories](#), [and what they are referring to](#), [approving or banning companies](#), [search logs](#) or [support with knowledge database](#).

[Google Modules](#)- [\[source code\]](#)

A iGoogle gadget directory released months before Google released their own. The site receives about 25 gadget submissions per day and has over 1700 del.icio.us bookmarks. Google Modules was co-produced with my friend [Philipp Lenssen](#).

A python rewrite of Google Modules is available [open source](#) for education purposes.

[Wikitrivia](#) - [\[source code\]](#)

Random automatic quizzes using wikipedia. An interesting experiment relased before the word mashup was first coined and which received over 50,000 hits on the first three days it launched. WikiTrivia is now [open soure](#).

COURSES

Relevant machine learning courses I took at Cornell:

COMS 478: Machine Learning with Golan Yona (A-)
COMS 482: Theory of Algorithms with Jon Kleinberg (A)
COMS 578: Empirical Method in Machine Learning with Rich Caruana (A+)
COMS 678: Advanced Topics in Machine Learning with Thorsten Joachims (A)
COMS 790: Independent Research with Rich Caruana (A+)
COMS 778: Topics in Machine Learning with Rich Caruana (A-)
COMS 750: Evolutionary Computation and Design Automation with Hod Lipson (A-)

Linear Algebra with Anil Nerode (A+), Multivariable Calculus (A), Groups and Geometry with Birgit Spohr (A+), Mathematical Physics with Kusse (A-)

PUBLICATIONS Marti A. Hearst, Anna Divoli, Harendra Guturu, Alex Ksikes, Preslav Nakov, Michael A. Wooldridge, and Jerry Ye. *BioText Search Engine: beyond abstract search*, [Bioinformatics](#), 2007
[\[pdf\]](#)

Chloé-Agathe Azencott, Alex Ksikes, S. Joshua Swamidass, Jonathan Chen, Liva Ralaivola, and Pierre Baldi. *One- to Four- Dimensional Kernels for Small Molecules and Predictive Regression of Physical, Chemical, and Biological Properties*, *Journal of Chemical Information and Modeling* (JCIM), 2006.
[\[html\]](#)

Rich Caruana, Alex Niculescu, Geoff Crew and Alex Ksikes, *Ensemble Selection from Libraries of Models*, Proceedings of the 21st International Conference on Machine Learning (ICML), 2004.
[\[ps\]](#) [\[pdf\]](#)

SOURCE CODE **Implementation of some Machine Learning algorithms with source code ...**

A java implementation of a [neural net](#). Solving the 8 queens problem with a [genetic algorithm](#) (from CS478). Implementation of the [k-means algorithm](#) using MDL for model selection. An experiment with [evolving artificial neural networks](#). An implementation of the [Gibbs sampling algorithm](#) to detect motifs in sequences (from CS478). Source code of [Shotgun](#) aka Ensemble Selection. An implementation of [greedy agglomerative clustering](#) with basic graphing. Quick and dirty [k-nearest neighbor](#).

Transcripts and references available upon request