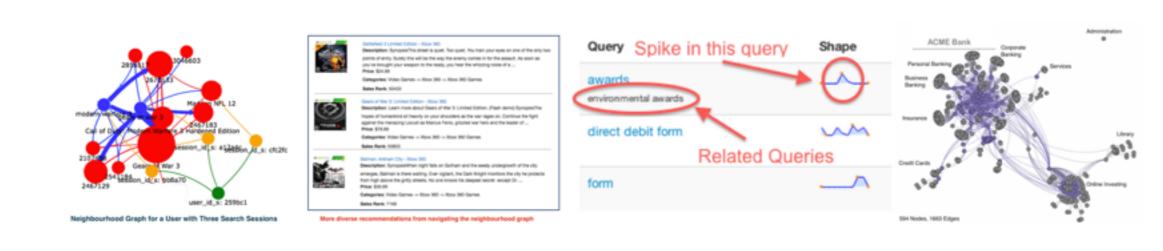
Combining ideas from the research and open source worlds to create new products & systems.



Summary (Bio)

I have designed and developed a 100 million page web crawler, recommender systems, a program for detecting "spikes" and related queries in search engine logs and a text miner for definitions from web pages.

I have previously worked for Funnelback, a search technology company with blue chip clients like Skype, Rolls Royce and Cambridge University. Prior to that I was a Research Engineer at Australia's national research agency, the CSIRO.

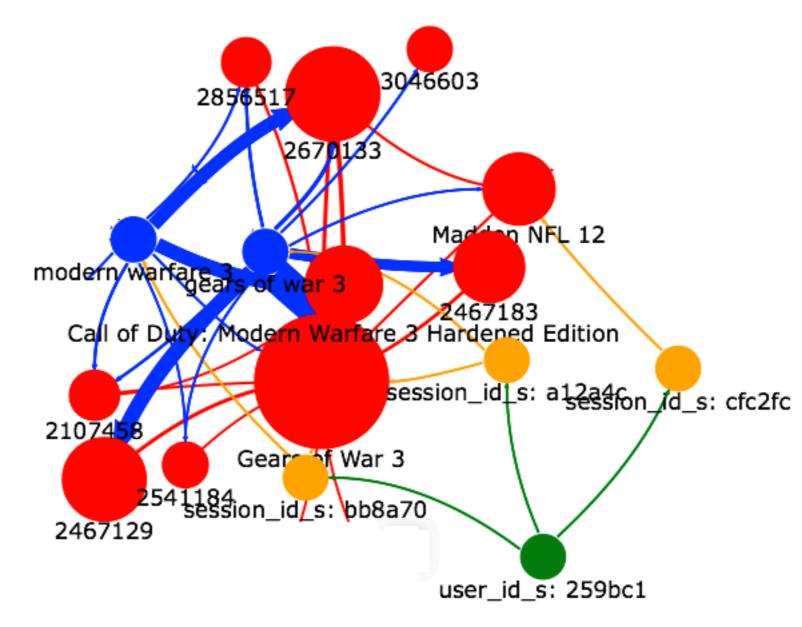
Portfolio

All of the systems or features shown in this portfolio were self-initiated, designed and implemented by myself, unless otherwise stated. The back-end systems were implemented in Java, while the front-ends were done using JavaScript.

Event Mining

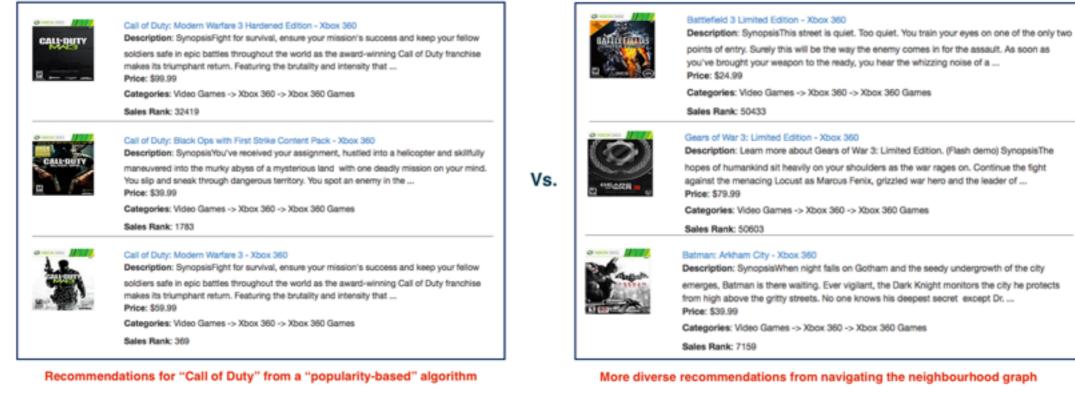
I designed and developed the EventMiner system for mining relationships from "event" logs for Lucidworks in California. The program generates a co-occurrence matrix from the raw event logs and then produces a graph from the matrix which encodes the relationships between users, sessions, queries and documents. The graph is encoded and stored in the Solr search engine, with a full-featured graph API layer also being provided.

The service can navigate the "neighbourhood graph" for an input item:



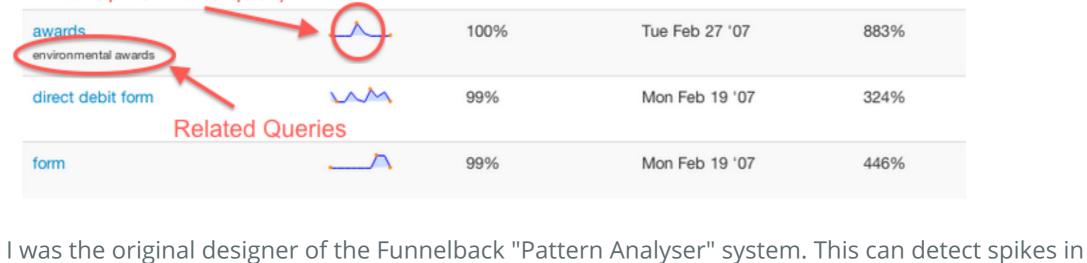
Neighbourhood Graph for a User with Three Search Sessions

to generate more diverse and high-quality recommendations:



Query Spike in this query Shape

Query Spikes and Related Queries



Confidence

Peak

Increase

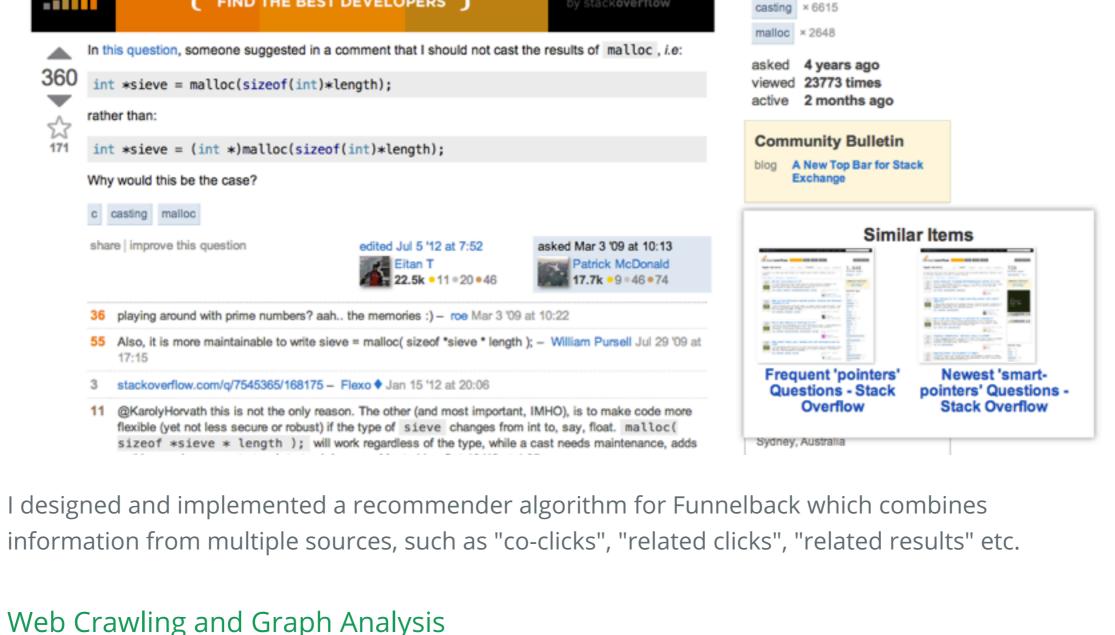
a news story). I also designed and implemented a system to automatically detect related queries, based on their click patterns. This can be used to group queries which have no terms in common but have the same semantics. Recommender Systems

the time series query log data, where these spikes are usually caused by some external event (e.g.

Do I cast the result of malloc?

tagged USE STACK OVERFLOW TO CAREERS 2.0 c × 118745 FIND THE BEST DEVELOPERS

Example of recommended "similar items" overlaid over a StackOverflow question:



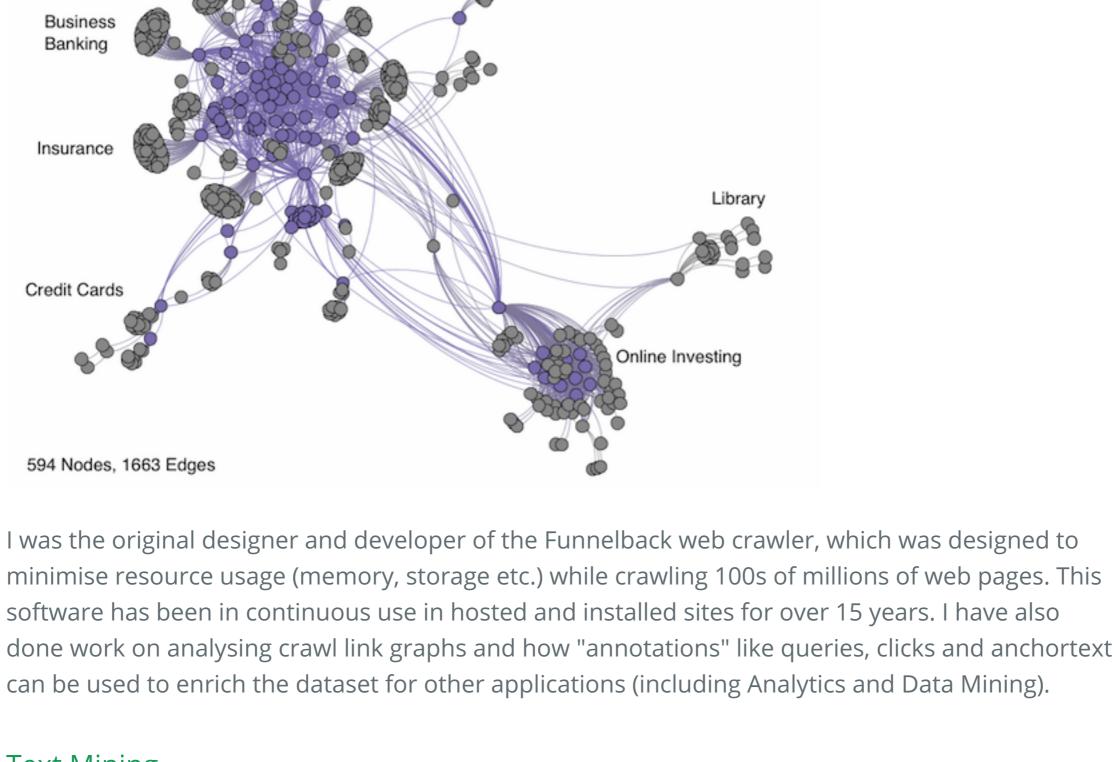
Example visualization of a web crawl:

Services

Administration

ACME Bank Corporate Banking

Personal Bankin



Text Mining Example of extraction of definitions from text: 1 - 4 of 4 search results for John Doe John Doe is a Distinguished Professor in the Department of Computer Science at the University of

I am the creator of a text mining system which uses syntactic patterns to extract definitions for

phrases and acronyms. When those terms are used as a query the definition can be displayed at

Key Skills

the top of the search results, with a link back to the page that contained the original definition.

Analytics

Massachusetts, Amherst...

- Information Retrieval & Search • Data Mining
 - Distributed Systems NoSQL Relational Databases
 - Project Management Java

• API Design

- Python JavaScript • Solr and Elasticsearch
- Spring, Spring Boot Vue.js

Automated Testing

- Education
- M.Sc. in Information Retrieval, Dublin City University (1996 1998)
- B.Sc. in Computer Applications (1st class honours), DCU (1992 1996)
- Contact

- Email: francis.crimmins @ gmail.com • Twitter: https://twitter.com/franciscrimmins
- LinkedIn: https://www.linkedin.com/in/franciscrimmins • GitHub: http://franciscrimmins.github.io/
- Disclaimer

This is a personal site and does not necessarily reflect the views or opinions of any of my previous employers or organisations I have done contracted development for. Details on example systems shown above refer to publicly available information from the official documentation sites for companies such as Lucidworks and Funnelback.

This page was generated by GitHub Pages using the Cayman theme by Jason Long.