

## Social Information Discovery

Search with a little help from your friends.

Barry Smyth  
CLARITY: Centre for Sensor Web Technologies  
University College Dublin

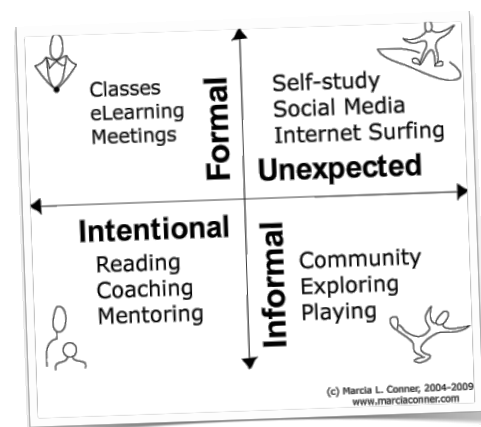


## Learning Evolution ...

Informal

Mobile

Social



Discovery

Collaboration

Curation



Web Search ?

## Motivations

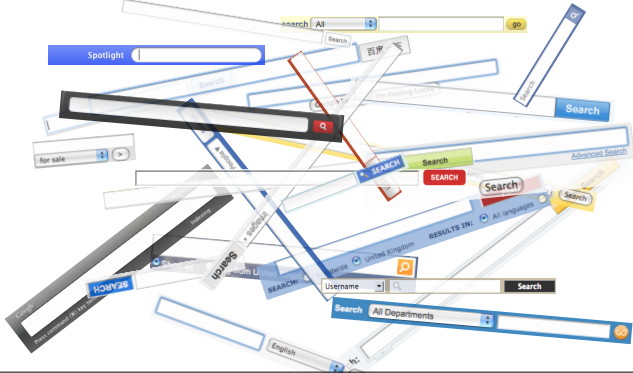
Informal learners need to be better supported during information discovery (web search/surfing)

Collaboration needs to be accommodated within our information discovery toolset.

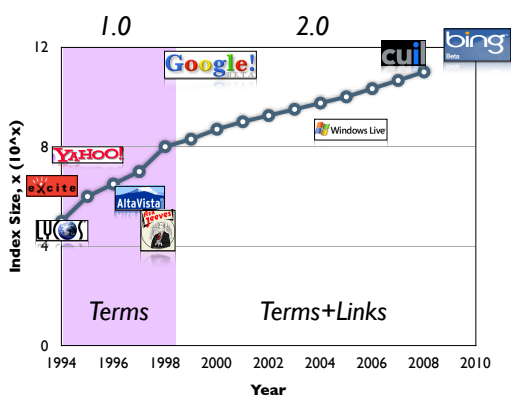
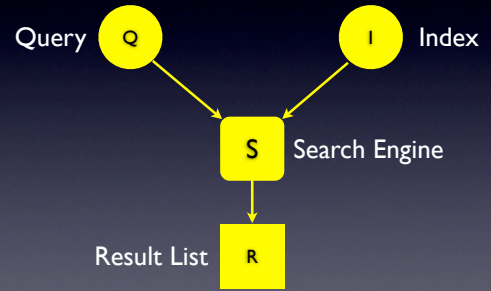
**BUT ...**

Mainstream search engines fail to provide this type of support infrastructure.

# The Web's Killer App?



# Web Search 101



Billions of queries per day!

# Overview

The State of Web Search - Key Challenges

Potential Solutions - Context in Web Search

Towards Social Web Search - HeyStaks

# Challenges

Vague Queries

The Vocabulary Problem

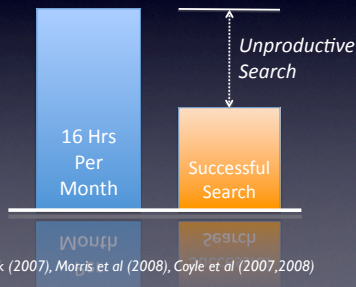
One-Size-Fits-All

Content Farming & SEO

# 50% Search Failure Rate

25% of searches ⇒ click to back button!

Average query size (2-3 terms) is insufficient to guarantee effective search engine retrieval



\* iProspect (2006), Jansen & Spink (2007), Morris et al (2008), Coyle et al (2007,2008)

# The Vocabulary Gap



Does Not  
One Size Fits All

A screenshot of a Google search result for the query 'umap'. The search bar shows 'umap' and the search button is visible. Below the search bar, there are several search results. The top result is 'UMAP - University Mobility in Asia and the Pacific' with a brief description and a link to 'www.umap.org'. Below it is 'Advanced Flash Components' and 'Umap (AS 3.0)'. The page also includes a 'Web Show options...' link and a 'Search' button.



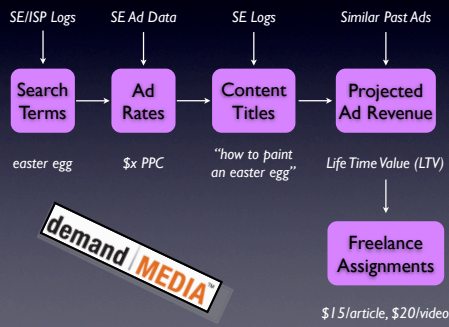
A screenshot of a search result for 'UMAP 2009'. The title is 'UMAP 2009 | UMAP2009'. The main text reads: 'UMAP 2009. The biennial conference series User Modeling (UM, 1986-2007) and Adaptive Hypermedia and Adaptive Web-Based Systems (AH, 2000-2008) have been ...'. Below the text are several links: 'Workshops', 'Travel Information', 'Registration', 'Important Dates', 'Accommodation', 'Program Committee', 'Reasons to Submit', and 'What's New'. At the bottom, there is a link for 'More results from fbk.eu'.

# Content Farms, SEO, Gaming

A screenshot of a Google search result for the query 'tune up a car'. The search bar shows 'tune up a car' and the search button is visible. Below the search bar, there are several search results. The top result is 'How to Tune Up a Car' from Lethow.com. Below it are 'Basic Car Tune-Up Video Series' and 'How Do I Tune Up My Car?'. The page also includes a 'Web Show options...' link and a 'Search' button. There are two annotations: a red box labeled 'demand MEDIA' pointing to the top result, and a red box labeled 'Read Write Web' pointing to a video result.

Focused SEO to promote commissioned content. (≈1m items / month)

# The DemandMedia Model



# Vague Queries

## The Vocabulary Problem

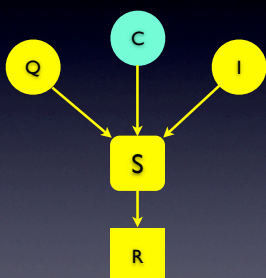
## One-Size-Fits-All

## Content Farming & SEO

# Web Search is changing...

# Improving search by better understanding user needs and search context ...

# Context in Search

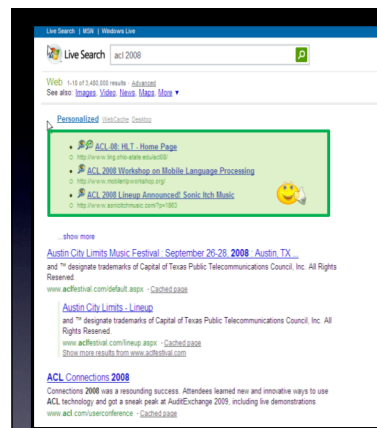


**User Context**  
Preferences, usage history, profiles

**Document Context**  
Meta-data, content features

**Task Context**  
Current activity, location etc.

**Social Context**  
Leveraging the social graph.



PSearch (Teevan et al, SIGIR 2005)

Client-Side Profiling

Explicit, Content, Behaviour

Personalized Ranking

# User Context

## Google AdSense

The screenshot shows a search results page for 'Flooring Dublin'. The search results include several articles from 'HomeTIPS' such as 'Installing The Flooring Hardware', 'Cork Flooring', 'Limestone Flooring', 'Bamboo Flooring', 'Recycled Tiles for Flooring', 'Resilient Flooring', and 'Resilient Flooring Maintenance and Repair'. To the right, there are several Google AdSense ads for flooring services, including 'Flooring Dublin' from Tradesmen.ie, 'Amilco & Carpet Specialist', 'TerraMai Flooring', and 'Wooden Floors'.

## Document Context

The screenshot shows a news article on smh.com.au titled 'Qantas emergency landing'. A red circle highlights the 'Qantas' logo on the aircraft. Another red circle highlights an advertisement for 'Fly To Australia Cheap' with the text 'Save up to 85% on Australia Flights With Our Discounted Airfares!'. The article text mentions 'Qantas Airways Boeing 747-400 passenger plane parked at the Henry Aquino International Airport showing its damaged right wing fuselage'.

## Task Context

**Activity/Task Context**  
E.g. writing a talk, planning a trip, shopping, etc.

**Location-Based Search**  
Google on the iPhone ...

**Embedded Web Search**  
Watson, (Budzik & Hammond, IUI 2000), RemembranceAgent (Rhodes, IEEE Trans. Comp 2003)

The screenshot shows an iPhone search interface with the query 'pizza'. The search results include 'Pizza Hut', 'PizzaHut.com - Menu', and 'PizzaHut.com/Menu.aspx'. The interface also shows a keyboard and a search button.

## IntelliZap

In appearance the Jaguar is often confused with the Leopard - both cats, depending on the species have a similar brownish/yellow marked with dark rosettes. The Jaguar can be distinguished by the shapes within the large muscular body and a s (see below) are comm often confusingly label also applied to black

The screenshot shows the IntelliZap search interface. The search query is 'jaguar'. The results include a list of related sites: Zapper, IntelliZap, Reference, News, Finance, Shopping, Movie info (movie title), General Recipe Archives, and Hotmail Inbox. The interface also shows a search button and a keyboard.

Context → Query Augmentation  
Finkelstein wt al. WWW 2001



1 in 4

... searches are for something the searcher has already found during a previous search session\*.

2 in 3

... searches are for something that a searcher's friends or colleagues have recently found\*.

Search should be more personal & collaborative!

\* Morris et al (2008), Teevan et al (2007), Smyth et al (2004,2006,2008)

# Web Search as a Social Activity



## Search Engines

Google, Bing, Yahoo, ...



### Searching

>800m users

### Page Index

~20 Billion Pages

### Queries

500m queries/day

### Relevance

Terms, Links, PageRank

## Social Networks

FaceBook, MySpace, Twitter, ...



### Sharing

>400m users

### Social Graph

~50 Trillion Relationships

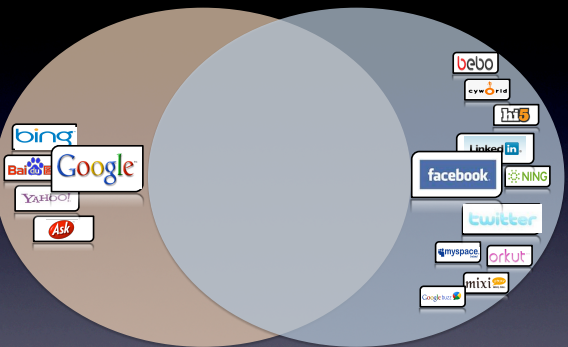
### Communities

750m million shares/day

### Reputation

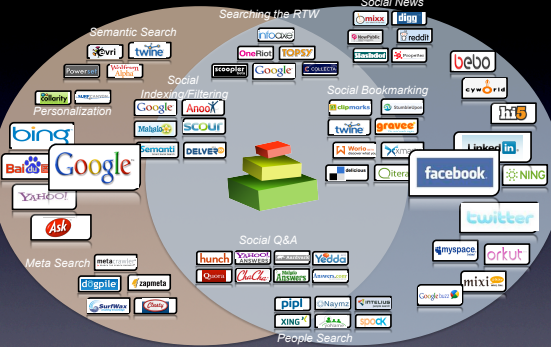
Social Rank?

Searching ← → Sharing



Queries ← → Communities

Searching ← → Sharing

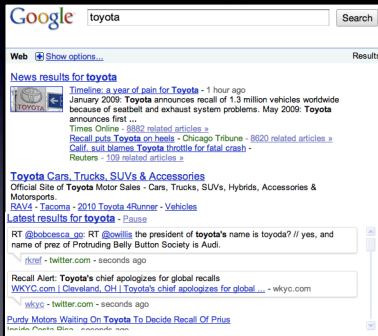


Queries ← → Communities



# Flavours of Social Search...

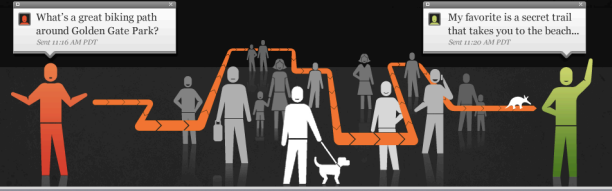
# Search & the Real-Time Web



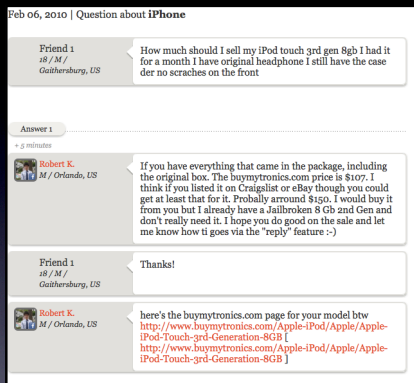
} Live Twitter Feed

# Exploiting the Social Graph

1. Send Aardvark a question
2. Aardvark finds the perfect person to answer
3. Get their response in a few minutes



www.vark.com – Social Q&A



A Conversational Thread on Aardvark

# Searching Social Content

# Harnessing the Social Graph

...

# Collaboration in Search

90% of people have engaged in some form of collaboration during Web search.

87% of people have exhibited "back-seat searching."

86% of people go on to share results with others.

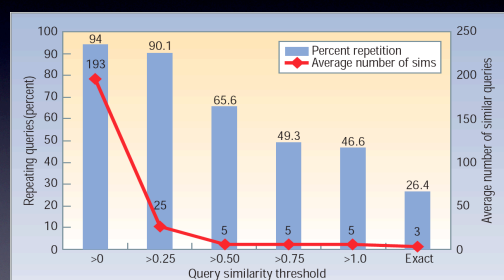
25%-40% of the time we are re-searching for things we have previously found.

66% of the time we are looking for something that a friends or colleague has recently found.

Morris et al., 2008

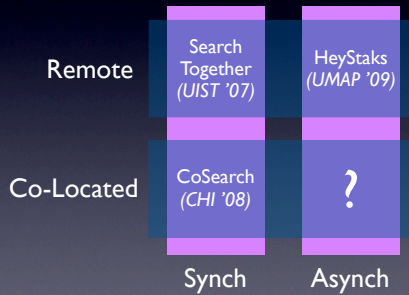
Smyth et al., 2006  
Teevan et al., 2007

# Repetition & Regularity in Communities ...



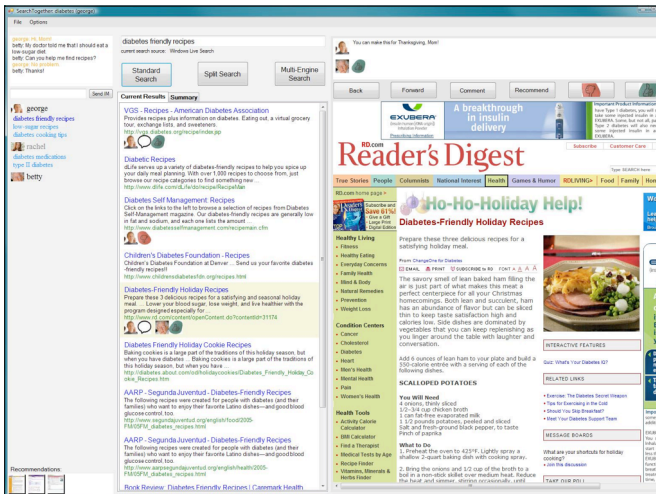
Smyth et al, UMUI 2004

# Collaborative IR



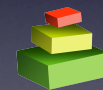
# SearchTogether

Morris et al (2008)



# HeyStaks

A Case-Study in Social Search



# Motivating HeyStaks

## Web Search, Shared!

Harness the collaborative nature of Web Search by providing integrated support for the sharing of search experiences.

## User Control

Support the searcher by providing fine-grained control over collaboration features and facilities.

## Integrate with Mainstream Search Engines

Users want to search as normal, using their favourite search engines, while, at the same time, benefiting from collaboration.

# HeyStaks: A Search Utility



## Create Staks

Users can easily create *Search Staks* (public/private) as a way to capture search activities.

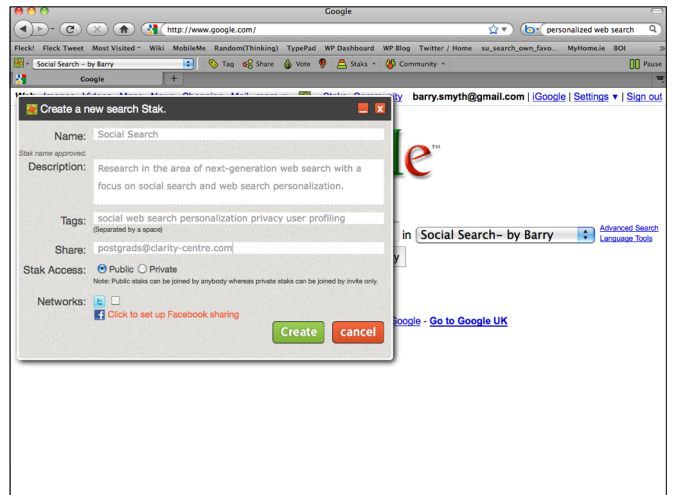
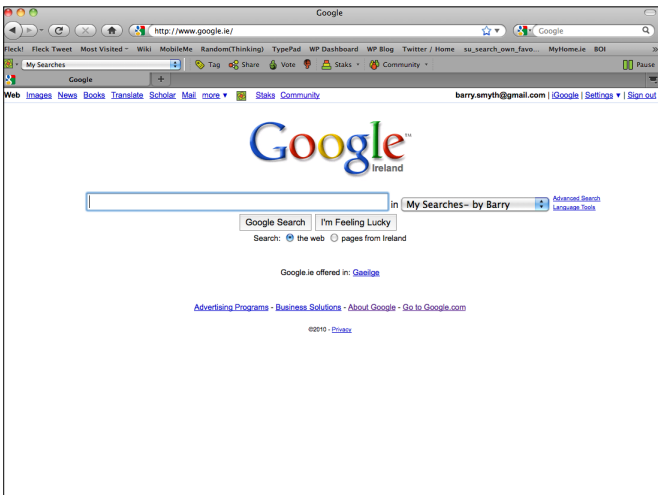
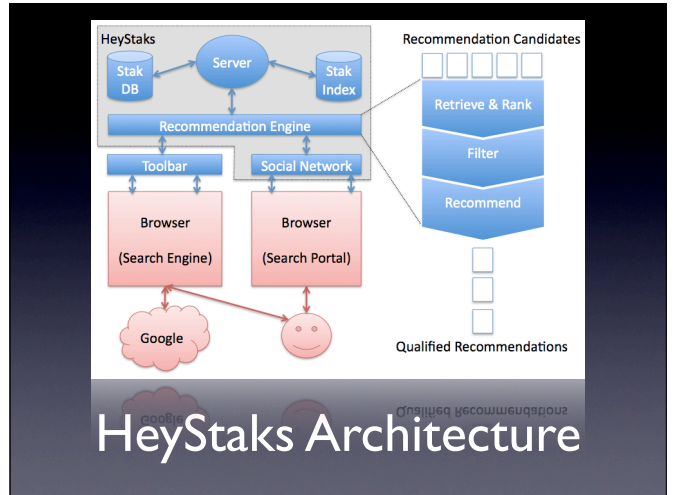
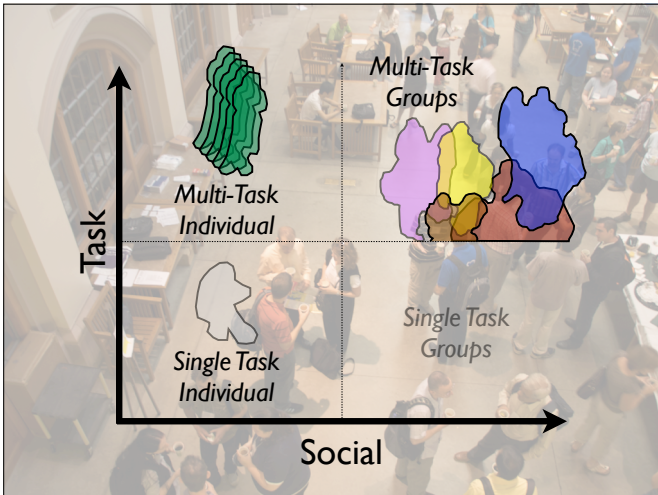
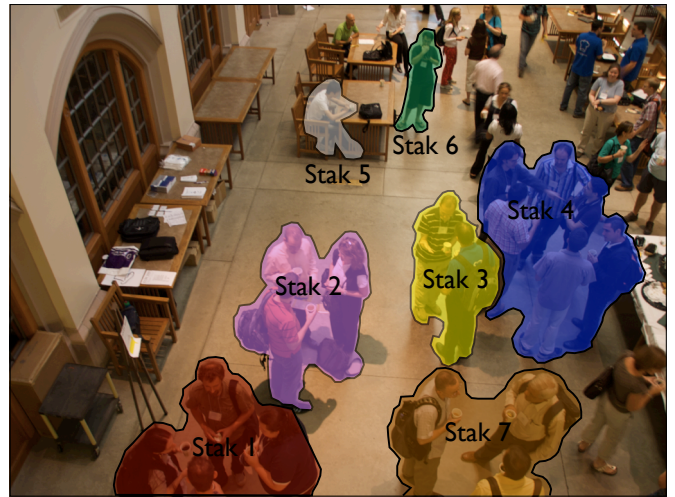
## Share Knowledge

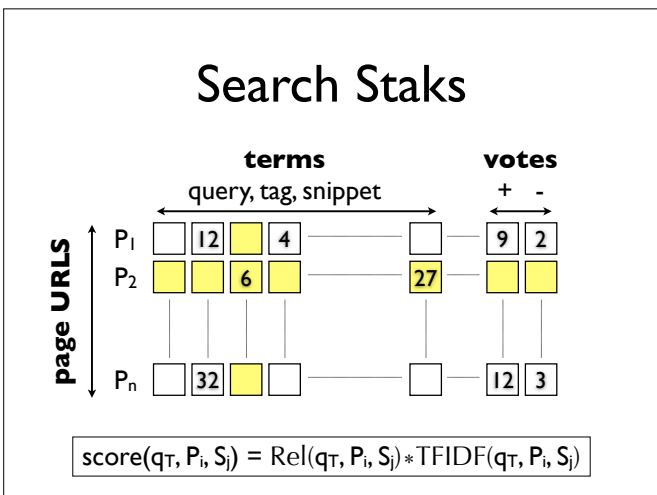
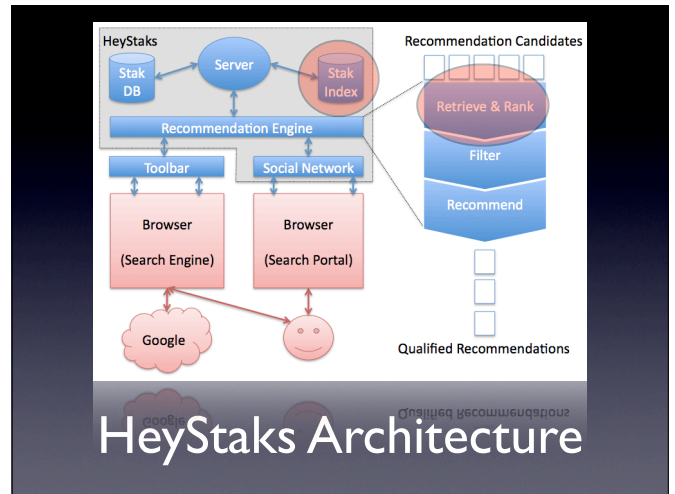
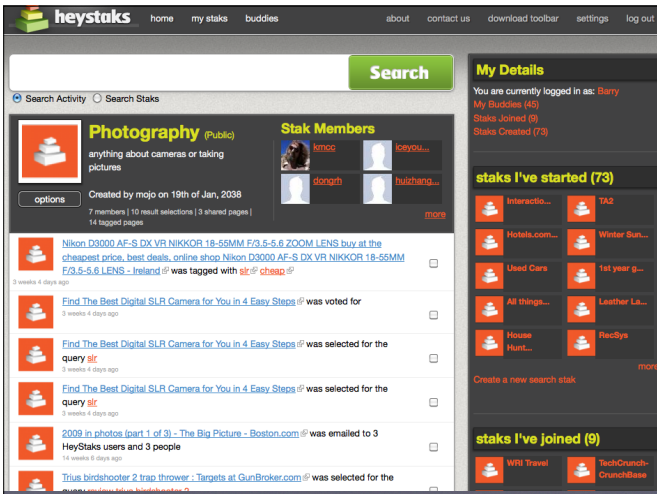
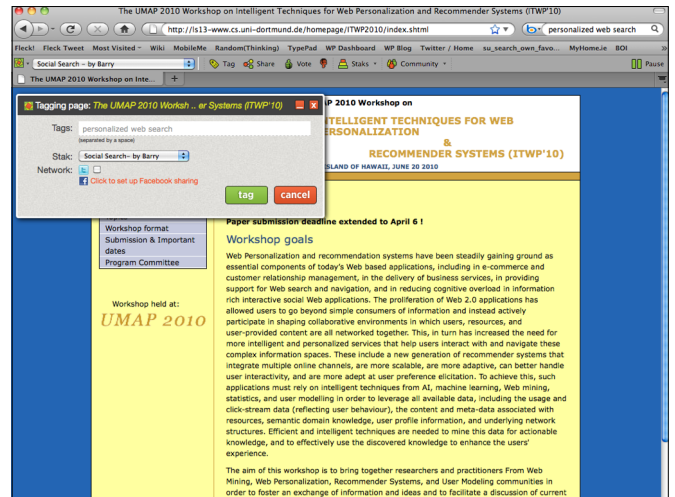
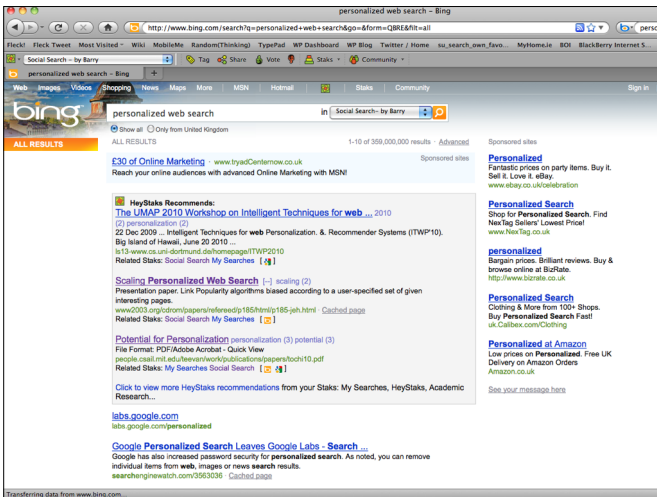
Share Staks with friends and others to grow community/task-based search expertise.

## Search & Promote

As users search within a Stak(s), relevant results are promoted and enhanced.







# The Social Life of Search

# Initial Evaluation

## HeyStaks Beta Trial

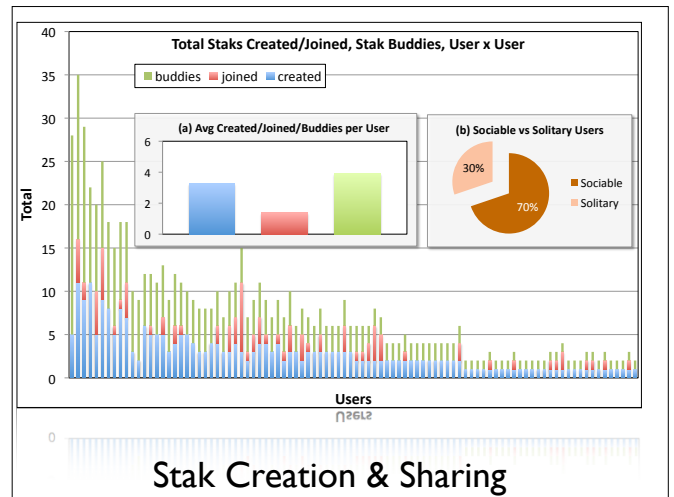
Focus on 95 early, active HeyStaks-Beta users who registered with HeyStaks during the period October 2008 - January 2009.

## Stak Creation/Sharing

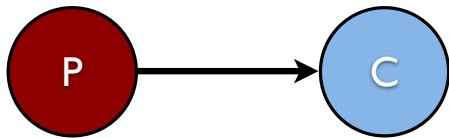
Do users take the time to create and share search staks (and search experiences)?

## Collaboration Effects

Do searchers benefit from the effects of search collaboration in general, and stak promotions in particular?

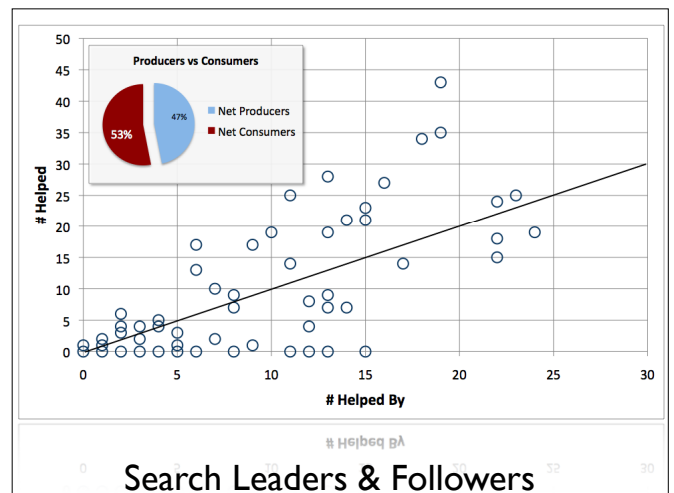
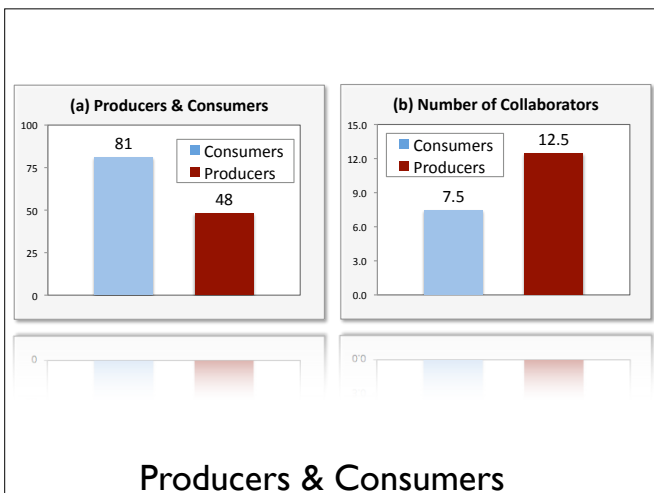
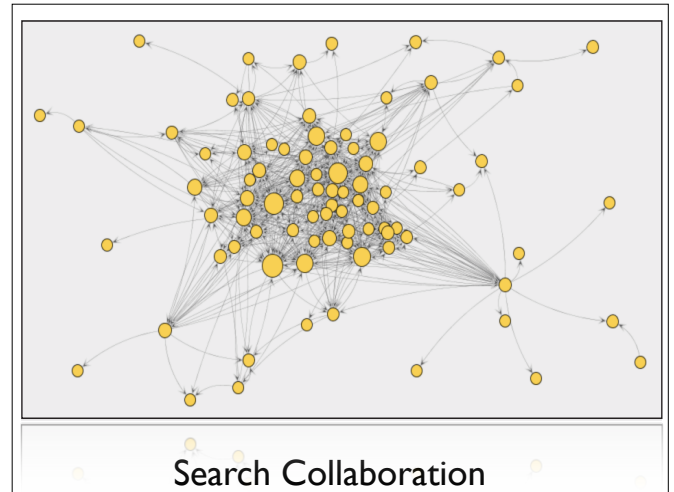


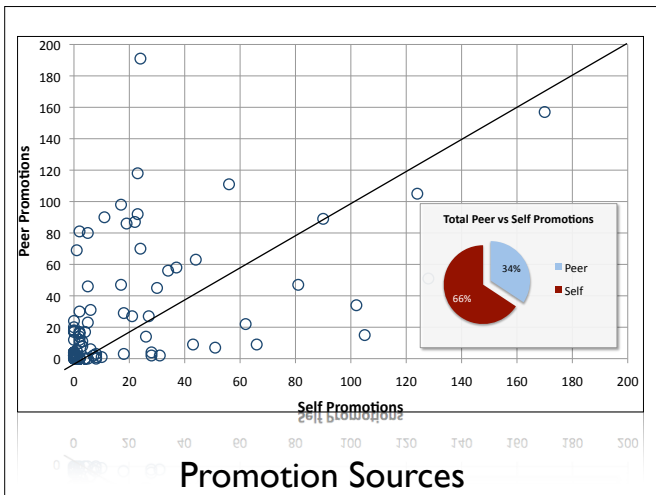
# Producers & Consumers



## Basic Unit of Collaboration

Searcher *C* selects a promotion previously selected by *P*.





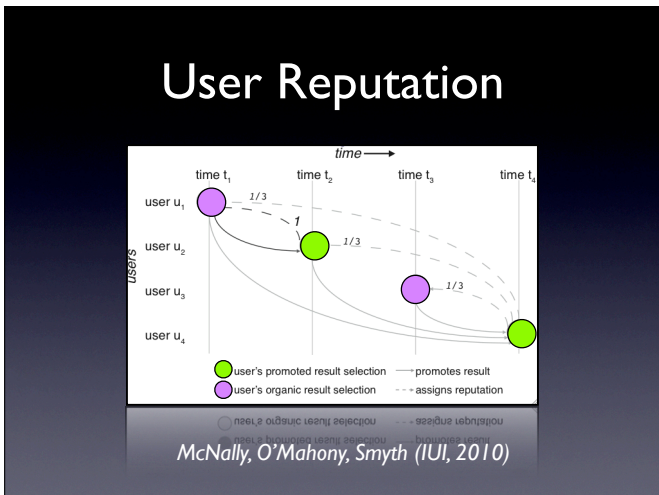
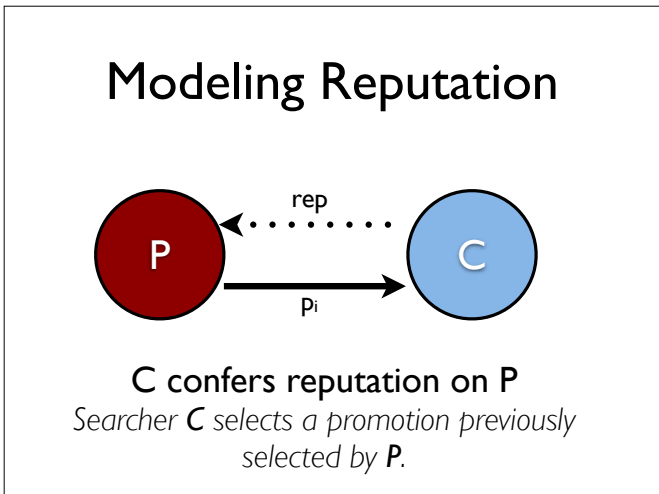
Users create & share staks.  
 Collaboration commonplace.  
 Users benefit from peers.

## Reputation

All searchers are not created equal! Staks are likely composed of a mixture of novice and expert searchers.

Can we identify the best searchers? Overall or at stak-level?

Can we use this *reputation* information to further influence recommendation?



## Reputation Model

**Producer Reputation**

$$rep(p_i, t) = rep(p_i, t - 1) + 1/k$$

**Result Reputation**

$$repscore(r, t) = \max_{p_i \in \{p_1, \dots, p_k\}} (rep(p_i, t))$$

**Reputation Ranking**

$$rankscore(r, q_t, p_1, \dots, p_k, t) = w \times repscore(r, t) + (1 - w) \times relscore(q_t, r)$$

# Initial Evaluation

## HeyStaks Reputation Trial

64 undergraduate students participated in a general-knowledge quiz using HeyStaks to guide their searches.

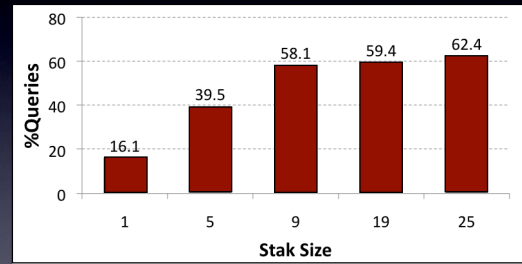
## Multiple Stak Sizes

Users were segregated into different stak sizes (1,5,9,19,25) to analyse the relationship between stak size and performance.

## Ground-Truth Based Performance Analysis

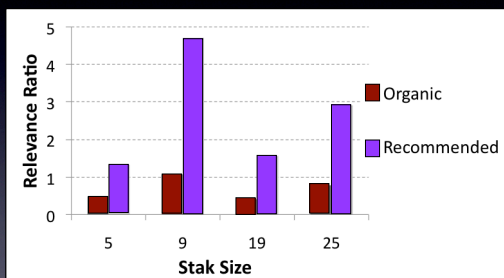
Fixed Q&A facilitated a definitive analysis of the relevance of organic and promoted results.

# Query Coverage



Percentage of queries receiving promotions.

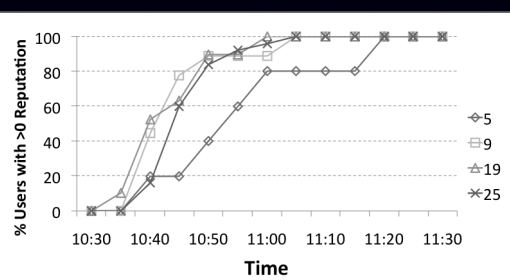
# Organic vs Promoted



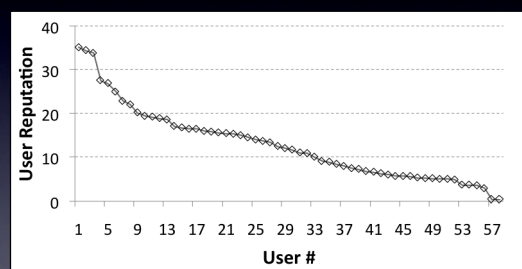
Relative relevance of organic and promoted results.

# Reputation Analysis

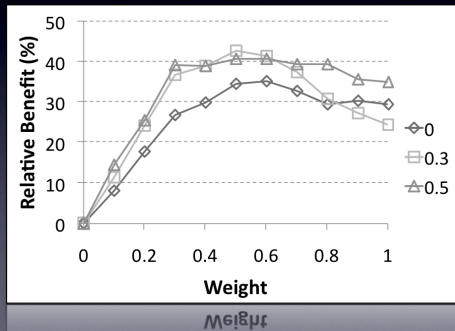
# Reputation x Time



# Final User Reputation



## Relative Benefit



## Conclusions

Web search and information discovery as a key part of informal learning.

The conservative world of Web Search is changing!

Collaboration in Web Search  $\Rightarrow$  Harnessing the Social Graph.

From relevance to reputation  $\Rightarrow$  Improved click-thru rates.

The screenshot shows the heystaks website interface. At the top, there's a navigation bar with 'about', 'contact us', 'download toolbar', and 'sign in'. The main content area features the heading 'ORGANISE SHARE & DISCOVER WITH OUR TOOLBAR'. Below this, a 'HOW IT WORKS' section is illustrated with three numbered steps: 1. 'You search & browse the web as normal (with the heystaks toolbar installed)', 2. 'We help you organise all the good stuff into staks (staks are groups of pages sorted by topic)', and 3. 'You can share these staks with friends & collaborate as you search.' A 'Create Stak' button is visible at the bottom of the main content area. The website URL 'www.heystaks.com' is displayed at the bottom.

## Lessons Learned

### Mainstream Web Search Integration

There is little value in developing competing Web search offerings; users want to search as normal using their favourite search engine (Google, Yahoo, Bing, ...)

### Personalization vs User Experience

An improved user experience can translate into much greater user-takeup than incremental improvements in personalization.