Search with a Little Help from Your Friends
Making Web Search more Collaborative

Barry Smyth
CLARITY: Centre for Sensor Web Technologies
University College Dublin
The Web’s Killer App?
The Web’s Killer App?
lycos.com (circa 1996)
Excite.com (circa 1996)
webcrawler.com (circa 1997)
“the verb” (Beta, circa 1998)
Google.com (circa 1999)
ask.com (circa 2000)
directhit.com (2000)
Approx. 1 billion queries per day!
Web Search 101

Query (Q) ➔ Search Engine (S) ➔ Result List (R) ➔ Index (I)
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

One Size Fits All

Does Not

^
UMAP - University Mobility in Asia and the Pacific
UMAP Special Workshop "Student Mobility for Language & Culture Studies", 15-17 August 2008, Ramkhamhaeng University, Bangkok, Thailand...
www.umap.org/ - Cached - Similar

UMAP - University Mobility in Asia and the Pacific -
Facilitate academic and research mobility by providing platforms for networking of interests among UMAP participating universities on, joint degree program...
www.umap.org/en/home/index.php - Cached - Similar
More results from www.umap.org »

Advanced Flash Components
UMap is the Universal ActionScript 3.0 Mapping API. The scalable component model allows you to create rich maps with interactive data layers. ...
www.afcomponents.com/components/umd_as3/ - Cached - Similar

UMap (AS 3.0)
In this tutorial we will describe the main principles of using UMap Routing API. We will also walkthrough a basic example for Flash. ...
www.afcomponents.com/tutorials/umd_as3/ - Cached - Similar
More results from www.afcomponents.com »
UMAP 2009 | UMAP2009
UMAP 2009. The biennial conference series User Modeling (UM, 1986-2007) and Adaptive Hypermedia and Adaptive Web-Based Systems (AH, 2000-2008) have been ...

umap09.fbk.eu/ - Cached - Similar

- Workshops
- Travel Information
- Registration
- Important Dates
- Accommodation
- Program Committee
- Reasons to Submit
- What's New
- More results from fbk.eu »
Focused SEO to promote commissioned-content. (≈1m items / month)
The DemandMedia Model

- SE/ISP Logs → Search Terms: easter egg
- SE Ad Data → Ad Rates: $x PPC
- SE Logs → Content Titles: “how to paint an easter egg”
- Similar Past Ads → Projected Ad Revenue → Life Time Value (LTV) → Freelance Assignments: $15/article, $20/video
Black Hat SEO

JC Penny Content Farming
New York Times, Feb 2011

Large-Scale Link Farming
2000+ sites linking to JCP for terms like “black dresses”, “casual dresses”, etc.

Significant Benefits to JCP
Millions of inbound visitors during the holiday season!
The Google Response

Major Algorithmic Change
Reducing the ranking of low-quality sites, impacting 12% of queries.

Spam-blocking Extension
Google Chrome extension to allow users to block low-quality sites from result-lists.

Major Impact on Many Sites
... including some surprises (SlideShare, Technorati).

Tuesday 3 May 2011
Vague Queries

The Vocabulary Problem

One-Size-Fits-All

Content Farming & SEO

Tuesday 3 May 2011
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.
User Context

Personalizing Web Search
Adapting search responses according to the needs/interests of the searcher.

Modeling User Interests
ODP Categories, query histories, result selections, etc.

Adaptation Strategies
Re-ranking, query modification, etc.

See also Chirita et al (SIGIR 2005), Shen et al (CIKM 2005), Teevan et al (TOCHI, 2010)
PSearch (Teevan et al, SIGIR 2005)

Client-Side Profiling

Explicit, Content, Behaviour

Personalized Ranking

User Context
Potential for Personalization

Teevan et al (SIGIR 2008)

How well can a single result-list satisfy a group of users?

Not every query benefits from personalization.

Predicting the PfP of search queries (click-based measures)
Qantas emergency landing

Qantas Airways Boeing 747-400 passenger plane parked at the Ninoy Aquino International Airport showing its damaged right wing fuselage. Photo: AP
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
In appearance the Jaguar is often confused with the Leopard - both cats, depending on species have a similar brownish/yellowish coat. However, the Jaguar is distinctively marked with dark rosettes or irregular shapes within the larger spots. It is more stocky and muscular body and a smaller head. Black Jaguars (see below) are common in some parts of the range and are often confusingly labeled as the same which is also applied to black
Web Search as a Solitary Activity

Tuesday 3 May 2011
Recovery vs Discovery

<table>
<thead>
<tr>
<th></th>
<th>Repeat Click</th>
<th>New Click</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeat Query</td>
<td>29%</td>
<td>4%</td>
</tr>
<tr>
<td>New Query</td>
<td>10%</td>
<td>57%</td>
</tr>
</tbody>
</table>

Teevan et al, 2007
1 in 4

2 in 3

... searches are for something that a searchers friends or colleagues have recently found*.

Search should be more personal & collaborative!

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?
Flavours of Social Search...
Search & the Real-Time Web

Live Twitter Feed
Exploiting the Social Graph

1. Send Aardvark a question
   What’s a great biking path around Golden Gate Park?  
   Sent 11:16 AM PDT

2. Aardvark finds the perfect person to answer
   My favorite is a secret trail that takes you to the beach...  
   Sent 11:20 AM PDT

3. Get their response in a few minutes

www.vark.com – Social Q&A
Feb 06, 2010 | Question about iPhone

Friend 1
18 / M / Gaithersburg, US

How much should I sell my iPod touch 3rd gen 8gb I had it for a month I have original headphone I still have the case der no scraches on the front

Answer 1

+ 5 minutes

Robert K.
M / Orlando, US

If you have everything that came in the package, including the original box. The buymytronics.com price is $107. I think if you listed it on Craigslist or eBay though you could get at least that for it. Probably arround $150. I would buy it from you but I already have a Jailbroken 8 Gb 2nd Gen and don't really need it. I hope you do good on the sale and let me know how ti goes via the "reply" feature :-(

Friend 1
18 / M / Gaithersburg, US

Thanks!

Robert K.
M / Orlando, US

here's the buymytronics.com page for your model btw
http://www.buymytronics.com/Apple-iPod/Apple/Apple-iPod-Touch-3rd-Generation-8GB
http://www.buymytronics.com/Apple-iPod/Apple/Apple-iPod-Touch-3rd-Generation-8GB
Key Features

Social Indexing
People vs Documents: User-Topic / User-User Relationships

Question Classification
Scored topic list (classification-based approach)

Query Routing
An aspect model routes questions to potential answerers.

Answer Ranking
Ranked candidate answerers based on topic, expertise, availability.

Damon Horowitz, Sepandar D. Kamvar: The anatomy of a large-scale social search engine. WWW 2010: 431-440

Tuesday 3 May 2011
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

Tuesday 3 May 2011
Repetition & Regularity in Communities...
Collaborative IR

Remote

Co-Located

Search Together
(UIST '07)

HeyStaks
(UMAP '09)

CoSearch
(CHI '08)

Synch

Asynch

Tuesday 3 May 2011
SearchTogether
Diabetes-Friendly Holiday Recipes

Prepare these three delicious recipes for a satisfying holiday meal.

From ChangeOne for Diabetes

The savory smell of lean baked ham filling the air is just part of what makes this meat a perfect centerpiece for all your Christmas homecomings. Both lean and succulent, ham has an abundance of flavor but can be sliced thin to keep taste satisfaction high and calories low. Side dishes are dominated by vegetables that you can keep replenishing as you linger around the table with laughter and conversation.

Add 6 ounces of lean ham to your plate and build a 550-calorie entree with a serving of each of the following dishes.

SCALLOPED POTATOES

You Will Need
4 onions, thinly sliced
1/2–3/4 cup chicken broth
1 can fat-free evaporated milk
1 1/2 pounds potatoes, peeled and sliced
Salt and fresh-ground black pepper, to taste
Pinch of paprika

What to Do
1. Preheat the oven to 425°F. Lightly spray a shallow 2-qt baking dish with cooking spray.
2. Bring the onions and 1/2 cup of the broth to a boil in a non-stick skillet over medium heat. Reduce the heat and simmer, stirring occasionally, until
done.

INTERACTIVE FEATURES

Quiz: What’s Your Diabetes IQ?

RELATED LINKS

Exercise: The Diabetes Secret Weapon
Tips for Exercising in the Cold
Should You Skip Breakfast?
Meet Your Diabetes Support Team

MESSAGE BOARD

What are your shortcuts for holiday cooking?

Join the discussion

TAKING THE PULSE
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.
HeyStaks Architecture

Tuesday 3 May 2011
Create a new search Stak.

Name: Social Search

Description: Research in the area of next-generation web search with a focus on social search and web search personalization.

Tags: social web search personalization privacy user profiling

Share: postgrads@clarity-centre.com

Stak Access: Public

Note: Public staks can be joined by anybody whereas private staks can be joined by invite only.

Networks: Facebook

Click to set up Facebook sharing

Create  cancel
Paper submission deadline extended to April 6!

Workshop goals

Web Personalization and recommendation systems have been steadily gaining ground as essential components of today’s Web based applications, including in e-commerce and customer relationship management, in the delivery of business services, in providing support for Web search and navigation, and in reducing cognitive overload in information rich interactive social Web applications. The proliferation of Web 2.0 applications has allowed users to go beyond simple consumers of information and instead actively participate in shaping collaborative environments in which users, resources, and user-provided content are all networked together. This, in turn has increased the need for more intelligent and personalized services that help users interact with and navigate these complex information spaces. These include a new generation of recommender systems that integrate multiple online channels, are more scalable, are more adaptive, can better handle user interactivity, and are more adept at user preference elicitation. To achieve this, such applications must rely on intelligent techniques from AI, machine learning, Web mining, statistics, and user modelling in order to leverage all available data, including the usage and click-stream data (reflecting user behaviour), the content and meta-data associated with resources, semantic domain knowledge, user profile information, and underlying network structures. Efficient and intelligent techniques are needed to mine this data for actionable knowledge, and to effectively use the discovered knowledge to enhance the users' experience.

The aim of this workshop is to bring together researchers and practitioners from Web Mining, Web Personalization, Recommender Systems, and User Modeling communities in order to foster an exchange of information and ideas and to facilitate a discussion of current
HeyStaks Architecture

HeyStaks

Stak DB

Server

Stak Index

Recommendation Engine

Toolbar

Social Network

Browser
(Search Engine)

Browser
(Search Portal)

Google

Recommendation Candidates

Retrieve & Rank

Filter

Recommend

Qualified Recommendations
Search Staks

\[ \text{score}(q_T, P_i, S_j) = \text{Rel}(q_T, P_i, S_j) \times \text{TFIDF}(q_T, P_i, S_j) \]
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?
Total Staks Created/Joined, Stak Buddies, User x User

(a) Avg Created/Joined/Buddies per User

(b) Sociable vs Solitary Users

30% Sociable
70% Solitary

Stak Creation & Sharing

Tuesday 3 May 2011
Producers & Consumers

Basic Unit of Collaboration

Searcher $C$ selects a promotion previously selected by $P$. 
Search Collaboration
(a) Producers & Consumers

(b) Number of Collaborators

Producers & Consumers

Tuesday 3 May 2011
Search Leaders & Followers

Tuesday 3 May 2011
Promotion Sources

Tuesday 3 May 2011
Users create & share stacks.

Collaboration commonplace.

Users benefit from peers.
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?
C confers reputation on P

Searcher C selects a promotion previously selected by P.
User Reputation

McNally, O’Mahony, Smyth (IUI, 2010)
Reputation Model

**Producer Reputation**

\[ \text{rep}(p_i, t) = \text{rep}(p_i, t - 1) + 1/k \]

**Result Reputation**

\[ \text{repscore}(r, t) = \max_{\forall p_i \in \{p_1, ..., p_k\}} (\text{rep}(p_i, t)) \]

**Reputation Ranking**

\[ \text{rankscore}(r, q_t, p_1, ..., p_k, t) = w \times \text{repscore}(r, t) + (1 - w) \times \text{relscore}(q_t, r) \]
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.

Tuesday 3 May 2011
Query Coverage

*Percentage of queries receiving promotions.*

Tuesday 3 May 2011
Organic vs Promoted

Relative relevance of organic and promoted results.
Reputation x Time
Final User Reputation

![Graph showing user reputation over user numbers](image-url)
ORGANISE
SHARE &
DISCOVER
WITH OUR TOOLBAR

You search & browse the web as normal.
(with the heystaks toolbar installed)

We help you organise all
the good stuff into staks.
(staks are groups of pages sorted by topic)

You can share these staks with
friends & collaborate as you search.

Try it out

www.heystaks.com

For example: "Winter Ski 2010", "Technology & Gadgets", "Old irish Literature Research", "Creating a Startup", "Social Search"
Conclusions

The conservative world of Web Search is changing!

Context in Web search \(\Rightarrow\) Adaptation.

Collaboration in Web Search \(\Rightarrow\) Harnessing the Social Graph.

From relevance to reputation \(\Rightarrow\) Improved click-thru rates.
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, ...)

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