Social Information Discovery
Search with a little help from your friends.

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

Informal
Mobile
Social

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**

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

**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

**Billions of queries per day!**
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

16 Hrs Per Month
Successful Search

Unproductive Search


The Vocabulary Gap

The average query size (2-3 terms) is insufficient to guarantee effective search engine retrieval. The image shows a graph comparing different search engines, indicating that

The Vocabulary Gap

One Size Fits All

Does Not

Content Farms, SEO, Gaming

Focused SEO to promote commissioned content (+1m items / month)
The DemandMedia Model

Web Search is changing...

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.

Vague Queries

The Vocabulary Problem

One-Size-Fits-All

Content Farming & SEO

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

PSearch (Teevan et al, SIGIR 2005)

Client-Side Profiling
- Explicit, Content, Behaviour

Personalized Ranking

User Context
Search should be more personal & collaborative!


Web Search as a Social Activity

Flavours of Social Search...
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.

---

Repetition & Regularity in Communities ...

Smyth et al, UMUAI 2004
Collaborative IR

Remote

Co-Located

SearchTogether
(Morris et al 2008)

Synch

Asynch

HeyStaks
(UMAP '09)

SearchTogether
(UIST '07)

CoSearch
(CHI '08)

HeyStaks
(UMAP '09)

SearchTogether
(UIST '07)

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

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?

Producers & Consumers

Basic Unit of Collaboration
Searcher C selects a promotion previously selected by P.

Search Collaboration

Producers & Consumers

(a) Producers & Consumers

(b) Number of Collaborators

(b) Producers & Consumers

Search Leaders & Followers

Producers vs Consumers

Net Producers

Net Consumers
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?

Modeling Reputation

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

\[ \text{resultscore}(r, i) = \text{max} \left( \text{max} \left( \text{resultscore}(r, j) \right) \right) \]

Reputation Ranking

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

<table>
<thead>
<tr>
<th>Stak Size</th>
<th>% Queries Receiving Promotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.1</td>
</tr>
<tr>
<td>5</td>
<td>39.5</td>
</tr>
<tr>
<td>9</td>
<td>58.1</td>
</tr>
<tr>
<td>19</td>
<td>59.4</td>
</tr>
<tr>
<td>25</td>
<td>62.4</td>
</tr>
</tbody>
</table>

Percentage of queries receiving promotions.

Organic vs Promoted

<table>
<thead>
<tr>
<th>Stak Size</th>
<th>Relevance Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Organic 0.5</td>
</tr>
<tr>
<td>9</td>
<td>Promoted 1.5</td>
</tr>
<tr>
<td>19</td>
<td>Organic 2.5</td>
</tr>
<tr>
<td>25</td>
<td>Recommended 3.5</td>
</tr>
</tbody>
</table>

Relative relevance of organic and promoted results.

Reputation Analysis

Reputation x Time

Final User Reputation
Relative Benefit

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

The conservative world of Web Search is changing!

Collaboration in Web Search ⇒ Harnessing the Social Graph.

From relevance to reputation ⇒ 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.