Improving Findability in the Enterprise
Lee Romero, Global Consulting Knowledge Management

• 4 years with Deloitte
• Leads SharePoint initiatives for Global Knowledge Services
• Portal program manager for Consulting
Introduction

• Many organizations are implementing an enterprise search solution and enterprise social networking
• They also hear frequently from users that “search doesn’t work”
• This session will:
  o Provide insights about how enterprise social networking tools can improve findability
  o Cover specific case studies of how social search is becoming more prevalent and how enterprise social networking tools can be used to specifically improve the findability of content within an enterprise search solution
  o Provide detailed insights on actionable analytics which can improve search results
  o Reveal the challenges you will face in understanding what your users are looking for and how you can work around that
  o Suggest how information architecture can be used to improve findability
Overview

• Enterprise search can make large collections of content available to more users and, in doing so, it makes it more and more difficult to ensure that users can find what they're looking for.
• Enterprise social networking tools bring more and more users together virtually and allow for simpler many-to-many conversations.
• At Deloitte, we have implemented an enterprise search solution.
• Our search covers over 1,000,000 separate targets, and that number is growing all of the time, both through organic growth of existing collections and addition of whole new collections to the search.
• We have also implemented an enterprise social networking tool that enables users across Deloitte to interact and post questions.
Social search
Social search

How our enterprise social networking tool has had the effect of improving findability of content by enabling our users to perform a social search – searching by asking questions of others.

From Patrick Lambe’s forward to Heather Hedden’s “The Accidental Taxonomist”:

This problem [of organization and retrieval of digital content collections] has largely been addressed on the internet through increasingly sophisticated search. Search can satisfy this need on the internet because it can exploit two unique features of the internet:

1. **Super availability of information** means that useful content can be found for almost any purpose

2. **Social clues** about useful content such as links and references can lift content that has been recognized as useful to the top of search results

This can be applied to the enterprise but in general here I will be talking about another kind of impact of social context to findability.
Recommend uses for Enterprise Social Networks

Share a link. “Here is a link to the latest Forrester Wave report on social networking.”

Ask a question. “Has anyone encountered this problem before, and if so, how was it solved?”

Find a resource. “Looking for a specialist in retirement benefits to help win a bid in Calgary.”

Answer a post. “Here are links to three relevant examples in the project database.”

Recognize a colleague. “Thanks to John Doe for hosting an excellent planning session today.”

Inform about your activities. “Will be in the Philadelphia office today; does anyone wish to meet?”

Suggest an idea. “Local office TV screens should display the global conversation stream.”
Ways to do a social search

• Post in the all company stream
• Post in a topic-specific group
• Post in the knowledge help desk group
All company

Jenny Wilson
We are looking for relevant experience and qualifications on developing a b2b customer engagement index - does anyone have anything in this space?
Like · Reply · Share · More · January 14 at 2:50pm from iPhone

Katherine Goodwin: here is a direct link to useful B2B quals from KX
https://kx.deloitteresources.com/_layouts/CES/PortalSearch.a...

https://kx.deloitteresources.com/_layouts...
kx.deloitteresources.com

January 14 at 3:49pm · Like · Reply · Share · More

Jenny Wilson: Thank you very much
January 14 at 4:02pm from iPhone · Like · Reply · Share
Topic-specific group

Stan Garfield › Deloitte Analytics: Official
Looking for experts in text analytics

Stan Garfield › Knowledge Management
Can you help @Greg Mitchell find people and/or documents on these topics?
• Text analytics
• Semantic Web
• iRing
• Natural Language Processing
• Machine learning
• Artificial Intelligence

View Full Conversation

Like · Reply · Share · More · November 16, 2012 at 12:52pm

David Steier: A few people attached and most of us at San Jose know a little about AI and machine learning.

cc: @Cindi Thompson, @Krishna Kumaraswamy, @Sandeep Kumar Sharma
November 16, 2012 at 5:14pm - Like · Reply · Share · More

Stan Garfield: Thanks, David!

cc: @Greg Mitchell
November 16, 2012 at 5:23pm - Like · Reply · Share · More

Greg Mitchell in reply to David Steier: Hey David - great - we
Knowledge help desk group

Joanne Cronin: Business Support eXchange (BSX)
Hi BSX - I am looking for information on ERP systems in retail clients. Preferably a breakdown along the lines of SAP vs Oracle vs Other ERP. Links to qual decks etc would also be appreciated
Like · Reply · Share · More · January 30 at 5:48am

Nidhi Rao: Hi Joanne,
The following documents may be useful to you:

Qualifications:
IT Directional Assessment, Pre ERP Business Case
1/4/1999 - 12/31/2012
Link:
https://quals.deloitteresources.com/search/ShowProject.asp?id=

CCR Process Optimization and SAP BPC Implementation Project
10/1/2009 - 4/1/2011
Link:
https://quals.deloitteresources.com/search/ShowProject.asp?id=

Core Merchandising and Warehousing System Selection
7/12/2010 - 12/17/2010 expand »

https://quals.deloitteresources.com/search/
quals.deloitteresources.com

January 31 at 7:02am · Like · Reply · Share · More
👍 Liked by Joanne Cronin.

Nidhi Rao: We have shared your request with Technology Advisory EMEA and Business Research Center.
January 31 at 7:09am · Like · Reply · Share · More
👍 Liked by Joanne Cronin.

Joanne Cronin: Thank you very much!
January 31 at 8:24am · Like · Reply · Share · More
Examples of queries and reaction to responses

• Any links to thought leadership articles on change management, talent and human capital in general?
  ○ Amazing! Thanks.

• I am looking for an expert in the field of Enterprise Social Networks or Web 2.0 technologies in a Knowledge Management context in general.
  ○ I already scheduled two additional interviews - thanks!

• I am working on a proposal. Do you know of any cases where we have been responsible for coordinating interviews with the media and/or identifying the proper spokesperson to speak with the media?
  ○ Thank you very much!
Crowdsourcing
Crowdsourcing

How the enterprise search program has used the enterprise social networking tool to crowdsource targeted improvements in search results for users for important and yet underperforming searches
A key process in improving any search solution is to be able to identify “problematic” search terms and then work to improve results for those terms.

The next section on Analytics will discuss different ways to identify these “problematic” search terms for further analysis.

Here, I will describe how we’ve been using our enterprise social networking tool to help improve results.
Identifying desirable targets

We have adopted a process to post on a regular basis (“Search term of the week”) with a specific search term.

We ask for our users to collectively identify and “vote” on specific resources that they consider the “best” for that term.

The most commonly voted resources are deemed as the best.

This also obviously feeds into the manual “best bets.”
Confirming desirable targets

A similar process we’ve used as an alternative is to pre-identify the resources and ask users to vote on those (or possibly suggest alternatives).

We adopted this as an alternative when some terms elicited very little feedback from users.

We had a knowledge manager do some research on these terms to identify potential targets (any existing best bets were also considered).
Crowdsourcing improvements – next steps

The remainder of the process, once “best” resources are identified for specific terms is part of our standard process:

- Use the resources as benchmarks for assessing relevancy of search (their position in the results)
- Focus content optimization efforts on those resources (ensuring they are indexed, they are well-titled, well-tagged, etc.)
- Consider changes in weightings within the search engine to improve the position of these results for the specified search terms
- Work to actively move these resources up in the results (generally)
Crowdsourcing a strategy

Another specific way we used our enterprise social networking tool was during a re-evaluation of our search strategy.

We elicited input from our end users as well as encouraging discussions around the strategy.

This approach engendered a significant conversation – some of which is shown here.

Many of the items and comments were eventually incorporated into our new strategy for 2015.
Analytics
Analytics

A review of specific analytics that can provide useful insights on your user behavior and support improvements to your enterprise search solution

Your search engine may provide some of these

You may need to integrate your web analytics tool with your search to obtain some of these

And some may require you directly use the log file generated from your search engine
Basic analytics

“Basic” analytics are ones that typically would be provided directly by your search engine or analytics tool. They include:

- Total searches
- Top N searches  
  - “Super search terms”
- Not found searches / error searches / failed searches
- Total distinct search terms
Some advanced analytics

These “advanced” analytics are typically ones that will require you to pull data from multiple sources, or to process the data to obtain more in-depth understanding:

- Percent of visits to your portal / website that include a search
- Pages of results viewed per search visit
- Pages of results viewed for search terms
- Words per search term
- Categories of search terms
Understanding results usage

Another important view of your search usage is to understand the usage of the results. Some metrics to consider here include:

• Most common result clicked for your top searches

• Most commonly clicked on result across all searches

• Percent of traffic for specific resources that is driven from search

• Percent of results clicked on from “best bets”

• Another view of “failed” searches

• Percent of results clicked by page (1st vs. 2nd, etc.)
What kind of search do you support?

To effect meaningful changes in your search solution, I recommend you also identify the power curve of your search log.

This curve shows the percent of all searches performed against the percent of distinct search terms used.

The steeper the curve, the more concentrated your searches.

The shallower the curve, the more exploratory your users are.

This chart shows two search solutions I work with.

* I’ll have more on this topic in the next section.
Key performance indicators

The metrics described here are not necessarily good as key performance indicators for your search solution. They can be useful to manage your search solution but will not mean much to others.

The best key performance indicator would go above and beyond your search solution – it would be to determine the percent of your users who get to the information they need, regardless of whether they use search, navigation or both.

That can be very hard to measure, but there are some indicators that you can use from your search analytics that can give you directional understanding.

- Percent of visits to your portal that use a search
- Average pages of results a user looks at in a search visit
- Percent of searches performed where at least one result is clicked
The Challenge of your Search Log
Understanding your search log

For enterprise search solutions\(^1\), the “80-20” rule is not true

The **language variability is very high** in a couple of ways (covered in the next few slides)

Yet having a good understanding of the language, frequency and commonality in your search log is critical to being able to make sustainable improvements to your search

\(^1\) This does not seem to apply equally to e-commerce solutions
Some facts about search terms

There’s an anecdote that goes something like, “80% of your searches are from 20% of your search terms”

- Equivalently, some will say that you can make significant impact by paying attention to a few of your most common terms (you can, but in limited ways)

Fact: in enterprise search solutions the curve can be much shallower:

This chart shows the inverted power curve for two different solutions I’m currently working with

In the second case, it takes 13% of terms to cover 50% of searches, and that is over 7000 distinct terms in a typical month!
Some facts about search terms: part 2

Another myth: a large percent of searches repeat over and over again

Fact: on enterprise search solutions, there is surprisingly little commonality month-to-month

Over a recent six month period, which saw a total of ~289K distinct search terms, only 11% of terms occurred in more than 1 month!

<table>
<thead>
<tr>
<th># of months</th>
<th># terms</th>
<th>% of searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>257665</td>
<td>89.2%</td>
</tr>
<tr>
<td>2</td>
<td>17994</td>
<td>6.2%</td>
</tr>
<tr>
<td>3</td>
<td>5790</td>
<td>2.0%</td>
</tr>
<tr>
<td>4</td>
<td>2900</td>
<td>1.0%</td>
</tr>
<tr>
<td>5</td>
<td>2019</td>
<td>0.7%</td>
</tr>
<tr>
<td>6</td>
<td>2340</td>
<td>0.8%</td>
</tr>
</tbody>
</table>
Some facts about search terms: part 3

Another myth: a good percentage of your search terms will repeat in sequential periods

Fact: There is much more churn even month-to-month than you might expect – in the period covered below, only about 13% of terms repeated from one month to the next (covering about 36% of searches)
What to do with your search log?

The summary of the previous slides:

• It is hard to understand a decent percentage of terms within a given time period (month)!
• If you could do that, the problem during the next time period isn’t that much easier!

The next sections describe a couple of research projects I’ve been working on to tackle these issues
Information architecture
Engage a findability expert

• Assess your current information architecture and environment
• Make concrete recommendations for improving findability
• Possible experts

Andrew Hinton  Peter Morville  Lou Rosenfeld
Map topics to targets

• Compile a list of all important topics to the organization
• Map these to the best existing sources and define
  o Best bets
  o Next best links
  o Synonyms (“see also”)
• Add to
  o Search
  o Navigation
  o Index
  o Other alternatives
• For topics lacking a current target
  o Create a stub page
  o Assign a page owner
  o Flesh out the page
Develop an information architecture

• Include all instances in your enterprise of
  o Intranets
  o Knowledge assets
  o Collaboration tools

• Design effective navigation to reach
  o Sites
  o Pages
  o Content
  o Discussions
  o Profiles
Provide multiple alternatives for users to find content

• Search
• Navigation
• Chat
• Voice
• Email
• Social search
• Indexes
• Guides
• Top 10 lists
• Most popular lists
• Multiple user categories
• User assigned tagging (social bookmarking)
• Facets
• Filters
Develop a user-centered design

• Include all instances in your enterprise of
  o Sites
  o Pages
  o Content
  o Discussions
  o Profiles
• Specify required information, such as
  o Date last modified
  o Contact name
  o “Like” button
  o Tagging
• Use this to provide a consistent user experience
Keep content current

• Develop a process to
  o Review
  o Refresh
  o Remove
• For all
  o Sites
  o Pages
  o Content
  o Discussions
  o Profiles
• Automate detection of
  o Broken links
  o Orphan sites
  o Sites not recently updated
Enable user participation

• Evolve from static intranet pages to dynamic social interaction
• Allow users to help manage the information ecosystem
  o Social bookmarking
  o Tagging
  o “Liking”
  o Commenting
  o Recommending content for reuse or removal
• Integrate
  o Intranets
  o Knowledge assets
  o Collaboration tools
Understanding your users’ information needs
Categorizing your users’ language

• Given the challenges previously laid out, using the search log to understand user needs seems very challenging

• Beyond the first several dozen terms, it is hard to understand what users are looking for
  – And those several dozen terms cover a vanishingly small percentage of all searches!

• However, it would be very useful to understand your users’ information needs if we could somehow understand the entirety of the search log

• How do we handle this? Categorize the search terms!
Categorizing your users’ language, p2

• So we need to categorize search terms to really be able to understand our users’ information needs.

• To do this, we face two challenges

  1. What categorization scheme should we use?
  2. How do we apply categorization in a repeatable, scalable and manageable way?

• For the first challenge, I would recommend you use your taxonomy

• The second challenge is a bit more difficult but is addressed later in this deck
Categories to use

- **Proposal**: Start with your own taxonomy and its vocabularies as the categories into which search terms are grouped

- Some searches will not fit into any of these categories, so you can anticipate the need to add further categories

- This exercise actually provides an excellent measurement tool for your taxonomy
  - You can quantitatively assess the percent of your users’ language that is classifiable with your taxonomy
  - A number you may wish to drive up over time
Automating categorization

• Now we turn to the hairier challenge – how can we automatically categorize search terms?

• To describe the problem, we have:
  1. A set of categories, which may be hierarchically related (most taxonomies are)
  2. A set of search terms, as entered by users, that need to be assigned to those categories
Automating categorization, p2

• The proposed solution is based on a couple of concepts:
  1. You can think of this categorization problem as search!
  2. You are taking each search term and searching in an index from which the set of results is the set of categories

• Question: What is the “body” of what you are searching?
• Answer: Previously-categorized search terms!

• Using this approach, you can consider the set of previously-categorized search terms as a corpus against which to search
  – You can apply all of the same heuristics to this search as any search:
    • Word matching (not string matching)
    • Stemming
    • Relevancy (word ordering, proximity, # of matches, etc.)
Automating categorization, p3

• Here’s a depiction of this solution
Automating categorization, p4: Bootstrapping

• This approach depends on matching to previously-categorized terms
  – Every time you categorize a new search term, you expand the set of categorized terms, enabling more matches in the future

• Bootstrapping: You can take the names of the categories (the terms in your taxonomy) as the first set of “categorized search terms”
  – This allows you to start with no search terms having been categorized at all
  – You run a first round of matching against the categories to find first-level matches
  – Take those that seem like “good” matches and pull those into the set of categorized search terms for a second iteration, etc.
  – Using this in initial testing resulted in 10% of distinct terms from a month being associated with at least one category

• Another aspect: Any manual categorization of common search terms will add to the success of categorization
Automating categorization, p5: Iterative
Automating categorization, p5: Iterative

• This approach also needs to be applied iteratively
  – You start with a set of categorized search terms and a new set of (uncategorized) search terms
  – You then apply this matching to the uncategorized search terms, getting a set of newly-categorized search terms (with some measure of probability of “correctness” of the match, i.e., relevancy)
  – You pull in the newly-categorized search terms and run the matching process again
  – Each time, as you expand the set of categorized search terms (from a previous match), you increase the possibility of more matches (in subsequent matches)
Automating categorization, p6: Iterative

• It will be beneficial to have a human review the set of matches for each iteration and determine if they are accurate enough
  – The measurement of relevancy is intended to do this but would likely only be partially successful

• Over time, using this process, you build up a larger and larger set of categorized search terms
  – This makes it more likely in future iterations that more terms will be categorizable
Automating categorization, p7: No matches

• There will always be search terms that do not get matched.
  – This may be because the terminology used does not match
  – This may be because there are no categories in the global taxonomy that
    would be useful for categorization

• The first issue would require a human to recognize the association
  (thus, categorizing the term and then enabling matches on future uses of
  that term)

• The second issue would require adding in new categories (not part of
  the global taxonomy)
  – And then categorizing the term into the newly-added category(ies)
Summary

• With this approach, we can take a set of search terms at any time and categorize them (partially) automatically
  – Over time, the accuracy of the matching will improve through human review-and-approval of matches
• We then are able to relate these information needs to a variety of other pieces of data:
  – Volume of content available to users – significant mismatches can highlight need for new content
  – Rating of content in these categories – can highlight that a particular area of interest has content but it isn’t quality content
  – Usage of content in these categories – could highlight navigational issues (e.g., when a category is much more highly represented in search than in downloads)
• This does not require directly working with end-users and is scalable
Additional benefits: Measuring your taxonomy

• As mentioned earlier, part of the challenge will be that there will be terms that do not match the starting categories (i.e., the global taxonomy)

• This actually highlights some valuable insight obtainable from this:
  – We can **identify gaps in our taxonomy** (terms requiring new categories)
  – We can identify areas of our taxonomy where we have many search terms associated with a taxonomy term and **consider if we need to either add or split search terms** in order to better match our users’ real language
  – We can identify **areas of the taxonomy that are of little use** in terms of the language used by our users
Additional benefits: What are key words for your users?

• Word counts – independent of term usage, what are the most common individual words?

• Word networks – we can understand the inter-relationships between individual words (which pairs occur commonly together, which words occur commonly for a given word)

These are not as much about information needs as about understanding the language users use (so this insight can help shape categorization)
Lessons learned
Lessons learned

Try alternatives
- Ask different groups
- Try different questions
- Offer the answer and ask people to poke holes in it

Regularly search
- Put yourself in your users’ shoes
- See the real results
- Work to improve key searches – invest in this!

Use outsiders
- Users
- Experts

Your search log
- Work on understanding the language your users are using
- Go beyond the “short head”
More resources

• For more about search and enterprise search, I recommend
  • “Search Patterns: Design for Discovery” by Peter Morville and Jeffery Callender
  • “Enterprise Search” by Martin White
  • “Search Analytics for your Site” by Louis Rosenfeld

• For more information about information architecture, I recommend
  • “Information Architecture for the World Wide Web: Designing Large-Scale Web Sites” by Peter Morville and Louis Rosenfeld

• Some web sites or blogs of potential interest
  • www.searchtools.com – edited by Avi Rappoport
  • blog.leeromero.org - my own writings on this topic (and many others around content and collaboration)