Explore. Discover. **Focus.**
Making Search Optimization Effective, Repeatable, and Scalable
07 November 2017
Deloitte is a professional services firm – with over 260,000 professionals worldwide and over $38US billion in annual revenue delivering services in audit & assurance, tax, consulting, financial advisory and risk advisory.

I have been working in knowledge management and enterprise search for almost 20 years
- Currently product owner of Deloitte's enterprise search solution and lead our virtual Search Optimization Center team.

Our search solution is available globally:
- We have over 100K unique users per month
- In total, they perform over 1M searches per month
- Searching among over 1M potential results.
Enterprise search user interface

1. Filters reflecting our taxonomies
2. "Answer cards" – curated results with a deeper structure than a single link – generally high level topics
3. Promoted results – curated results that are a single link – generally very specific
4. Organic results – the results derived from content
Organization around our search solution

- **End users (10,000s)**: Users of search and user-generated content
- **Content managers and site managers (100s)**: Responsible for tagging and titling of content
- **Search optimization Center (~3 dozen)**: Help guide search optimization efforts – the "group brain"
- **Knowledge Services team & Technical team (about a dozen)**: Administer search, monitoring, education, build systems, manage relevancy

© 2017. For information, contact Deloitte Touche Tohmatsu Limited.
Our approach to search optimization

**Targeted**

- "Looking for trouble" and addressing it
- Using many of the techniques described by Rosenfeld: Failure analysis, session analysis, pattern analysis and others
- Biggest focus is on failure analysis – to align with our KPIs

**Education and content governance**

- Teaching content publishers to do SEO on their own
- Broad quality measurements of content sources based largely on heuristics
Targeted optimization
Our approach to search optimization – Analytics!

The ability to effectively optimize your search experience has 3 key elements:

1. Data – Even simple, standard reports right from our search engine and web analytics tool were critical at one point

2. Data – We have continued to evolve and refine our metrics over time – sharpening our focus

3. Data – We continue to work toward fully-automated metrics that align with our KPIs

I highly recommend you use Lou Rosenfeld's book "Search Analytics for your Site" as a starting point
Search optimization – the "hamster wheel"

- Identify issue:
  - Poorly performing terms or
  - Specific request or
  - Site-specific analysis

- Identify key results
- Analyze user demographics
- Assess search results
- Optimize content / Search experience
- Test outcome
Targeted optimization
Our ongoing evolution to finding problems

Search Analytics for your Site does a great job describing approaches to solving problems – but, critically, does not address the challenge of identifying problems.

Factors we grapple with:

- >100K users / month
- >250K distinct search terms / month
- Fewer than 2% of those terms are used >25 times / month
- Fewer than 30% of terms are used month-over-month

So a big focus of our evolution has been to continually refine the answer to the question, *What do we pay attention to?*
Focusing our problem identification microscope – Phases 1 and 2

The start

Initially we only had standard web analytics – page views and visits – and the search term was not captured directly!

1. Our first attempts were to simply order by page views
   - That equated to a user viewing a page of results
   - This ensured we looked at most-used terms
   - We had to parse the term off of the URL for some time
   - Understanding problems was a manual effort to "replay" those search terms ourselves

2. In a second phase, we adopted the page views / visit ratio as another indicator (no new base metrics, though)
   - We realized this gave us some idea of how much effort users were putting into their search – how many pages of results does a user look at on average for a term?
   - At this same time we changed search engines and used some of the in-engine reporting
     - Proved hard to reconcile, though
Focusing our problem identification microscope – Phase 3
Simplifying reporting and defining "search"

Our next step forward included:

- Capturing the search term as its own variable
- Adding in a "search" event (fired whenever a set of results is retrieved)

The search event was especially important because we adopted an "infinite scroll" results page – there is no new page view to see the next set of results.

We could then more easily generate reports based on rank ordered by usage (number of searches) and then differentiating based on searches / visit (typical effort)
Focusing our problem identification microscope – Phase 4

Defining a "search session" and capturing clicks

The fourth step in our evolution was to define two new measures:

- Unique searches – Counts the first search in each user session for a term (what users normally think of a "search")
- Clicks – counting the clicks, aggregated by search term!

These represented a big step for us – we could now calculate new ratios (aggregated by term):

- Searches / unique search – much more precise effort estimate
- Clicks / search – a clickthrough or "success" rate
Focusing our problem identification microscope – Phase 4
Defining a "search session" and capturing clicks

Using these we could now define multiple thresholds for combinations of effort and success that identified "poorly performing terms"

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row Labels</td>
<td>Searches</td>
<td>Visits</td>
<td>Clicks</td>
<td>Unique Visitors</td>
<td>Search Visit Ratio</td>
<td>Click Visit Ratio</td>
<td>Promoted Result?</td>
<td>Super Search Term?</td>
<td>Categorization</td>
</tr>
<tr>
<td>2</td>
<td>&quot;deloitte digital&quot;</td>
<td>230</td>
<td>42</td>
<td>56</td>
<td>36</td>
<td>5.48</td>
<td>1.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>aviation</td>
<td>202</td>
<td>27</td>
<td>50</td>
<td>23</td>
<td>7.48</td>
<td>1.85 Promoted Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>pricing</td>
<td>176</td>
<td>33</td>
<td>48</td>
<td>29</td>
<td>5.33</td>
<td>1.45 Promoted Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>internal control</td>
<td>175</td>
<td>30</td>
<td>59</td>
<td>27</td>
<td>5.83</td>
<td>1.97 Promoted Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>value proposition</td>
<td>140</td>
<td>20</td>
<td>29</td>
<td>17</td>
<td>7.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>business process</td>
<td>135</td>
<td>18</td>
<td>24</td>
<td>15</td>
<td>7.50</td>
<td>1.33 Promoted Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>sap security</td>
<td>121</td>
<td>17</td>
<td>36</td>
<td>10</td>
<td>7.12</td>
<td>2.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>supermarket</td>
<td>114</td>
<td>10</td>
<td>17</td>
<td>6</td>
<td>11.40</td>
<td>1.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>bersin by deloitte</td>
<td>108</td>
<td>18</td>
<td>8</td>
<td>13</td>
<td>6.00</td>
<td>0.44 Promoted Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>methods and tools</td>
<td>107</td>
<td>17</td>
<td>7</td>
<td>13</td>
<td>6.29</td>
<td>0.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>reference architecture</td>
<td>106</td>
<td>15</td>
<td>26</td>
<td>12</td>
<td>7.07</td>
<td>1.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>&quot;industry print&quot;</td>
<td>102</td>
<td>19</td>
<td>24</td>
<td>15</td>
<td>5.37</td>
<td>1.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>job description</td>
<td>102</td>
<td>14</td>
<td>20</td>
<td>13</td>
<td>7.29</td>
<td>1.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Clicks / search still suffered from a "research" problem, though:

- We could miss problematic terms because of one person researching a topic (and clicking on a lot of results)
Focusing our problem identification microscope – Phase 5

Our current model

For fifth step forward (and current model) adds in another new metric:

- "First click" – counts the first click in any user's session for a particular term

This removes the "research" problem!

**We can now calculate:**

- Abandonment rate = 1 - (first clicks / unique searches)
- This is the ratio of search sessions with no click!
- We now use this measurement overall as a KPI and at the individual term level to identify problematic terms
Focusing our problem identification microscope – Phase 5

Our current model

Our current reports are automatically generated from our web analytics tool daily, weekly and monthly

- We also provide specialized reports by user demographic (line of business or geography) and make those available to targeted members of our Search Optimization Center team

We can usually identify problems very quickly with these – new topics and terms appear all of the time

We now regularly find ourselves "ahead" of content providers!
Education and governance
Education and awareness

Or, how to get those 100s of content publishers to help improve your search?

**Search education**

- Search optimization office hours
- Publishing and maintaining detailed guidelines for content publishers from all content sources – specific "how to's" – empower and educate them!

**Search awareness**

- Targeted sessions with subsets of content managers
- Regular portions ("commercials") in community events for our intranet publishers
- Reinforcing the message that it is our content publishers who are responsible for the quality of search
Content governance

To help reinforce the responsibilities of content publishers we recently implemented a "site health report" for intranet sites. The report is broader than search – it is an attempt to measure the quality of the management of intranet sites.

A site’s score is derived from three factors:

- **Usage**
- **Age**
- **Findability (built up from titling and tagging assessments)**

Just this year we have adopted a new guideline that sites can be excluded from enterprise search based on their health scores!
## Site health report - sample

<table>
<thead>
<tr>
<th>Site Collection</th>
<th>Name of the site</th>
<th>Number of Pages</th>
<th>Overall Grade</th>
<th>Age Grade</th>
<th>Page Usage Grade</th>
<th>Findability Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>C</td>
<td>1</td>
<td>2.20</td>
<td>3.00</td>
<td>3.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Caribbean and Bermuda</td>
<td>C</td>
<td>1</td>
<td>1.80</td>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Central Europe</td>
<td>A</td>
<td>5</td>
<td>2.76</td>
<td>3.80</td>
<td>1.40</td>
<td>2.40</td>
</tr>
<tr>
<td>Central Europe</td>
<td>L</td>
<td>24</td>
<td>2.17</td>
<td>2.08</td>
<td>1.50</td>
<td>2.58</td>
</tr>
<tr>
<td>Central Europe</td>
<td>T</td>
<td>38</td>
<td>3.11</td>
<td>4.74</td>
<td>1.63</td>
<td>2.21</td>
</tr>
<tr>
<td>Chile</td>
<td>S</td>
<td>4</td>
<td>2.90</td>
<td>4.00</td>
<td>1.00</td>
<td>2.75</td>
</tr>
<tr>
<td>CIS</td>
<td>W</td>
<td>2</td>
<td>3.40</td>
<td>4.00</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Consulting</td>
<td>A</td>
<td>13</td>
<td>4.75</td>
<td>5.00</td>
<td>4.54</td>
<td>4.62</td>
</tr>
<tr>
<td>Consulting</td>
<td>C</td>
<td>1</td>
<td>4.20</td>
<td>5.00</td>
<td>5.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Consulting</td>
<td>G</td>
<td>10</td>
<td>4.84</td>
<td>5.00</td>
<td>4.40</td>
<td>4.90</td>
</tr>
<tr>
<td>Consulting</td>
<td>Ir</td>
<td>1</td>
<td>4.20</td>
<td>5.00</td>
<td>5.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>A</td>
<td>11</td>
<td>2.38</td>
<td>3.00</td>
<td>1.00</td>
<td>2.45</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>C</td>
<td>1</td>
<td>1.80</td>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Deloitte Brasil</td>
<td>N</td>
<td>6</td>
<td>2.80</td>
<td>3.00</td>
<td>2.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Denmark</td>
<td>C</td>
<td>44</td>
<td>2.87</td>
<td>3.50</td>
<td>1.82</td>
<td>2.77</td>
</tr>
<tr>
<td>Denmark</td>
<td>F</td>
<td>61</td>
<td>2.65</td>
<td>3.23</td>
<td>1.46</td>
<td>2.67</td>
</tr>
<tr>
<td>Denmark</td>
<td>P</td>
<td>1</td>
<td>1.80</td>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>DrPubStageBefore</td>
<td>D</td>
<td>3</td>
<td>1.80</td>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>EMEA Region</td>
<td>E</td>
<td>1</td>
<td>4.20</td>
<td>5.00</td>
<td>5.00</td>
<td>3.00</td>
</tr>
<tr>
<td>EMEA Region</td>
<td>If</td>
<td>3</td>
<td>1.80</td>
<td>1.67</td>
<td>1.00</td>
<td>2.33</td>
</tr>
<tr>
<td>Global</td>
<td>C</td>
<td>60</td>
<td>3.36</td>
<td>3.37</td>
<td>2.10</td>
<td>3.98</td>
</tr>
<tr>
<td>Global</td>
<td>Ir</td>
<td>47</td>
<td>4.33</td>
<td>4.66</td>
<td>3.00</td>
<td>4.66</td>
</tr>
<tr>
<td>Global</td>
<td>N</td>
<td>21</td>
<td>3.80</td>
<td>3.57</td>
<td>4.90</td>
<td>3.48</td>
</tr>
<tr>
<td>Global</td>
<td>S</td>
<td>9</td>
<td>3.58</td>
<td>4.78</td>
<td>3.44</td>
<td>2.44</td>
</tr>
<tr>
<td>Global</td>
<td>U</td>
<td>2</td>
<td>4.20</td>
<td>4.00</td>
<td>3.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Global Audit &amp; Assurance</td>
<td>U</td>
<td>14</td>
<td>4.60</td>
<td>5.00</td>
<td>4.00</td>
<td>5.00</td>
</tr>
</tbody>
</table>

© 2017. For information, contact Deloitte Touche Tohmatsu Limited.
Next step with governance

In parallel with the adoption of the site health report we have also adopted new requirements for new content sources:

- Their content is subject to a quality review before they "get in the door" of enterprise search.

- The owners of the content source must provide a recurring report to help ensure quality expectations continue to be met.

These are too new to know the full effect and we are not yet applying these retroactively to all previously-integrated sources.
Summary
Summarizing

Key elements to improving your search:

**Data** – Data is critical to make decisions!

**Evolution** – We knew we did not have the best data available when we started - we focused on taking action with what we had and improving our "vision" over time!

**Education and awareness** – We ensure that content publishers understand the important role they play and we hold them to it!

**Content governance** – Garbage in, garbage out, so we work to keep the garbage out in the first place!