Search 101

August 23, 2019,
Maish Nichani (maish@pebbleroad.com)
Agenda

- About me
- About search
- 4 search stories
- What is enterprise search
- Elements of search experience
- Search strategy
- Search projects
What does enterprise search mean to you?
“Just make it work like Google”
Want search to work like magic
But starts like a black box not many understand
The journey
Enterprise search
Enterprise search leverages powerful stack of technologies to tame large volumes of heterogeneous information in order to offer users a fruitful search experience.
Elements of search experience
User
We use search to become<one of these>for a while

Assistant  Explorer  Analyst  Executive
How many pounds in a kg?

Do I need a visa to enter Myanmar?

What does this transaction code mean?
What do I advertise for Java developers?
Who sells good wooden furniture?
What are my employment benefits?
Which is the best phone to buy for dad?

Does that company really sell ‘fair-trade’ coffee?

How does the VW Golf compare with other cars in its class?
Executive

How has sales fared since the apology for cheat software in our cars?

How many cases were opened this year and how many of them closed?

Which areas in Singapore have the highest office rental yield?
How many pounds in a kg?
Which areas will give highest rental yield?
Which phone to buy for dad?
Infoseeking behaviours

Formulate query → Refine results → Find perfect match

Initial query

Results
Expertise

- Expert/Novice
  - Can start with using the right keywords

- Double expert
  - No problem.

- Double novice
  - Big problem.

- Novice/Expert
  - Can pick up by exploring the collection

*Designing the Search Experience*, Tony Russell-Rose and Tyler Tate
Persona

I want to...

Name of collection

Maximise or increase

Minimise or lower

So I can...
I want to watch a comedy movie after a hard day's work

Find the right movie quickly
Watch in good quality resolution

No buffering pauses
No other distractions

Unwind and relax

Netflix
Content
Content

Google search for chocolate chip cookie recipe.
Modelling for relevance
1. No metadata, no search
2. Bad metadata is worse than no metadata
### ATM collection

<table>
<thead>
<tr>
<th>#</th>
<th>Location</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Thomson Road</td>
<td>213 Thomson Road Singapore 432123</td>
</tr>
</tbody>
</table>

### ATM collection (optimised)

<table>
<thead>
<tr>
<th>#</th>
<th>Location</th>
<th>Address</th>
<th>System status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Thomson Road</td>
<td>213 Thomson Road Singapore 432123</td>
<td>Out of order</td>
</tr>
<tr>
<td>Metadata</td>
<td>Description</td>
<td>Values</td>
<td>Source</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------------</td>
<td>----------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>The region that the agency falls under.</td>
<td>North, South, East,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>West, Central</td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location zip code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact phone number</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Context
Context

Searching from Bukit Timah
Context

Who
Identity

Where
Location

What
Activity

When
Time

Towards a Better Understanding of Context and Context-Awareness,
<table>
<thead>
<tr>
<th>Persona</th>
<th>Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td></td>
</tr>
<tr>
<td>Environment</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td></td>
</tr>
</tbody>
</table>
Scenario

When Jack is in office and searching for a topic from the projects page, results from projects similar to the one he is working on right now are ranked higher.

If Jack is new to the company, in addition to the search results, results from the How-to section are ranked higher.
Interface
Interface

The beautiful box
Input
Features that allow the searcher to express what they are looking for (e.g., search bar)

Control
Features that help searchers to modify, refine, restrict or expand their Input (e.g., filters)

Informational
Features that provide results or information about results (e.g., snippets)

Personalisable
Features that relate specifically to searchers and their past interactions (e.g., bookmarks)
1-16 of over 100,000 results for "mobile phones"

LG
Sleek, new technology with touchless commands
LG G8 ThinQ with Alexa Hands-Free – Unlocked SMARTPHONE – 128 GB – Aurora

 NUU Mobile G3 Unlocked Cell Phone 64GB + 4GB RAM - 5.7" Android Smartphone - Sapphire Blue

$139.99
Eligible for Shipping to Singapore

NUU Mobile G3 Plus Unlocked Cell Phone 64GB + 4GB RAM 4000 mAh
Twitter Inc
TWTR as of Aug 21 4:00 PM ET | Markets are closed
42.80 • 0.51 | 1.21%
More at NASDAQ | Data by Xignite

Twitter Inc
NYSE: TWTR
42.80 USD +0.51 (1.21%) ↑
Closed: 21 Aug, 7:59 pm GMT+4 - Disclaimer
After hours 42.80 +0.080 (0.19%)

<table>
<thead>
<tr>
<th>1 day</th>
<th>5 days</th>
<th>1 month</th>
<th>6 months</th>
<th>YTD</th>
<th>1 year</th>
<th>5 years</th>
<th>Max</th>
</tr>
</thead>
</table>

| Open  | 42.83  |
| High  | 43.28  |
| Low   | 42.36  |
| Mkt cap | 33.06B |
| P/E ratio | 14.18 |
| Div yield | - |
| Prev close | 42.29 |
| 52-wk high | 43.48 |
| 52-wk low  | 26.19 |
How does the search engine know that the capital of India is Delhi?
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autocomplete</td>
<td>Autocomplete will attempt to complete the search term just as the searcher starts typing.</td>
</tr>
<tr>
<td>Autosuggest</td>
<td>Autosuggest will suggest resources found in the collection to the user while the user is still typing their search terms. It helps speed up retrieval.</td>
</tr>
<tr>
<td>Best bets</td>
<td>Best bets will elevate a document to very top of the list, ignoring any ranking algorithms. Can be useful for promoting certain pages on a website.</td>
</tr>
<tr>
<td>Context</td>
<td>Context plays a large part in improving relevancy and provides a personalised experience. Includes location, time, user profiles, weather, personal preferences, and other attributes to influence the search ranking algorithm for a custom result set tailored for the user.</td>
</tr>
<tr>
<td>Did you mean</td>
<td>A subset of spell checking, ‘did you mean’ can provide alternate terms to search for that closely matches the original term. Did you mean provides helpful clues for finding resources.</td>
</tr>
<tr>
<td>Faceted navigation</td>
<td>Facets act like filters to narrow down the results to find specific resources. This approach supports the exploration and discovery of information.</td>
</tr>
<tr>
<td>Field collapsing</td>
<td>When multiple documents are very similar in nature (eg, same movie, different format), consider collapsing the documents by a particular field, such as their title, to return greater variety in the search results.</td>
</tr>
<tr>
<td>Hit highlighting</td>
<td>Highlights query terms found in the resources. Helps users to quickly scan results to determine which ones are worth investigating further. Very useful for large documents.</td>
</tr>
<tr>
<td>Instant answers</td>
<td>Giving answers in the search results page or in autosuggest. Saves the user time from having to visit pages to get direct answers. This is very useful in mobile contexts.</td>
</tr>
<tr>
<td>Feature</td>
<td>Description</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Instant results</td>
<td>Similar to autosuggest, instant results will automatically update the search results page with new results when a character is typed into the search field. Saves time from having to type the full word or query.</td>
</tr>
<tr>
<td>Onboarding</td>
<td>Guides users through the search experience with smart tool tips. Helps novice users become more savvy with search.</td>
</tr>
<tr>
<td>Recommendations</td>
<td>Provide additional, similar documents based off of a certain criteria that can be customised.</td>
</tr>
<tr>
<td>Results bookmarking</td>
<td>Save a search snippet as a bookmark for easy retrieval at a later date.</td>
</tr>
<tr>
<td>Results grouping</td>
<td>Groups resources by specific categories in the search results. This helps to view different types of resources returned for the same term.</td>
</tr>
<tr>
<td>Results ranking</td>
<td>Providing options to rank the results by a specific criteria. Most common is by relevance score but others include dates, prices, ratings.</td>
</tr>
<tr>
<td>Rich snippets</td>
<td>Search snippets that present a rich set of relevant content to the user based. Often based on user persona (assistant, explorer, analyst, executive).</td>
</tr>
<tr>
<td>Search history</td>
<td>Automatically save the trail of searches made for easy retrieval at a later date.</td>
</tr>
<tr>
<td>Spell checking</td>
<td>Provides alternate spelling of words in a query to correct spelling mistakes. Helps to improve usability of the search experience. Very useful in mobile contexts.</td>
</tr>
<tr>
<td>Visualisations</td>
<td>View a bird's eye view of your collection in various chart types to see a bigger picture on the collection of information. Useful for the analyst user persona.</td>
</tr>
</tbody>
</table>
Technology
Technology

Dates
Duplicates
Entity extraction
Faceting
Multi-lingual
People search
Query management
Ranking
Sentiment analysis
Spell checking
Stemming and lemmatization
Stop words
Summarisation
Tokenization
...

Commercial

Open source
**Technology**

**Extraction tools**
- extracts raw data from the database, CMS, HTML, XMI, CSV files, and piped into the transformation layer

**Raw data**
- can be messy
- might have duplicates
- poorly formatted or no standard format
- doesn’t contain the right metadata relevant for search

**Transformation layer**
- data is sent through various modules that perform specific tasks (e.g., removing duplicate entries, standardizing formats)
- each object in resource can be enhanced with metadata by using auto-classifiers, named-entity recognition, and other NLP techniques
- the amount of transformation required will vary from one app to the next, depending on the condition of data and business goals

**Enriched data**
- data stored here is “clean” and modeled for search purposes

**Search engine**
- the search engine is configured and tuned to match the business objectives

**Frontend**
- frontend search framework

**Indexing**
- data from the enriched datastore is indexed into a search engine

**Analytics and personalisation**
- custom logs are generated
- logs used for analytics and fine-tuning search results
- logs are piped to a personalisation engine to provide a more personalised search experience
Technology

1. Strength
2. Scalability
3. Satisfaction - ease of use

You may find that open source search engines are equal to, if not better than, commercial ones.
All together

Search goal: Apply for paternity leave.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Example</th>
</tr>
</thead>
</table>
| 1. The result snippet shows the title and content around the keywords. | <a href="http://internal.yourcompany.com/policies/leave/">HR@YourCompany I Leave Policy</a>

You can find all policies related to leave in one place. How cool is that? |
| 2. The result snippet shows the title and matching subhead on paternity leave. | <a href="http://internal.yourcompany.com/policies/leave/">HR@YourCompany I Leave Policy</a>

Not only can you find all policies in one place, but you can also navigate to the policy you want right here. How cool is that?

Annual leave | Sick leave | Maternity leave | Paternity leave | Childcare leave | … |
| 3. The result snippet tells how many days of paternity leave the user is eligible for. | Hi John Malkovich, based on your records

You are eligible for <b>12 days</b> of paid paternity leave. [Apply for paternity leave]

Or find out about the <a href="conditions_of_leave">conditions of leave</a>.
Governance
Governance

“The impact of search on business performance depends more on the level of investment in a skilled team of people to support search than it does on the level of investment in search technology.”

Martin White, author of *Enterprise Search*
How is search performance measured?

- How many users are using the search box?
- What are the top \( xx \) places users begin their searches?
- What are the top \( xx \) search queries?
- What are the top \( xx \) pages reached through search?
- What are the queries with 0 results?
- What are the queries with 0 click-throughs on the SERP?
- How many times was search used immediately after the first SERP?
- What are the queries that result in more than 3 SERPs?
- How many users used facets or filters?
- How many users use “best bets”?
- ...
Planning a enterprise search project
The plan

AS-IS

A

TO-BE

B

Design activities

People, process, technology, governance
The process
<table>
<thead>
<tr>
<th>Phase</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discover the opportunity</td>
<td>Review target resource collection, Interview target users, Review search interfaces, Analyse search logs, Review search technologies, Review search governance</td>
</tr>
<tr>
<td>Define the desired experience</td>
<td>Identify dominant search personas, Identify core search tasks, Identify content models for relevance, Identify search interface features, Identify core technology requirements, Identify core measurement metrics, Define search governance requirements</td>
</tr>
<tr>
<td>Design a search prototype</td>
<td>Prepare resource collection for prototype, Design search interfaces, Deploy search technology for prototype, Design and test search prototype</td>
</tr>
<tr>
<td>Develop the new search service</td>
<td>Prepare actual resource collection, Implement search technologies, Develop search interfaces, Configure the search system, Test system with actual users</td>
</tr>
<tr>
<td>Govern the new experience</td>
<td>Analyse search clickstream logs, Analyse site search logs, Conduct search usability tests</td>
</tr>
</tbody>
</table>
Thank you