search
the lab
perspectives
Search is not just Google
Web
Images
Video
Translate
More

Brussels - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Brussels Cached
Brussels, officially the Brussels-Capital Region, is the de facto capital of the European Union (EU). Brussels

Bruxelles
bruxelles.around.com Cached
The City of Brussels is the largest municipality of the Brussels region, capital of Belgium, though the Brussels Capital Region

Bruxelles | Define Bruxelles at Dictionary.reference.com/browse/bruxelles Cached
Search is not just web search
Search is not just in a browser
Search is everywhere
Search - setting the stage

• 500 000 000 web searches every day
Search - setting the stage

- 500,000,000 web searches every day

- Half the world’s population uses web search
Search - setting the stage

- 500 000 000 web searches every day
- Half the world’s population uses web search
- Web search is trusted more than traditional media
Search - setting the stage

MEDIA SOURCES: SEARCH ENGINES NOW MOST TRUSTED
Trust in each source for general news and information (20-country global data)

<table>
<thead>
<tr>
<th>Year</th>
<th>Online Search Engines</th>
<th>Traditional Media</th>
<th>Hybrid Media</th>
<th>Social Media</th>
<th>Owned Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>67%</td>
<td>62%</td>
<td>52%</td>
<td>45%</td>
<td>43%</td>
</tr>
<tr>
<td>2013</td>
<td>63%</td>
<td>60%</td>
<td>50%</td>
<td>43%</td>
<td>42%</td>
</tr>
<tr>
<td>2014</td>
<td>64%</td>
<td>63%</td>
<td>53%</td>
<td>45%</td>
<td>44%</td>
</tr>
<tr>
<td>2015</td>
<td>65%</td>
<td>62%</td>
<td>53%</td>
<td>48%</td>
<td>47%</td>
</tr>
</tbody>
</table>

Millennials Are Even More Trusting Of Digital Media

<table>
<thead>
<tr>
<th>Source</th>
<th>Percentage</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online Search Engines</td>
<td>72% (+8)</td>
<td></td>
</tr>
<tr>
<td>Traditional Media</td>
<td>64% (+2)</td>
<td></td>
</tr>
<tr>
<td>Hybrid Media</td>
<td>63% (+10)</td>
<td></td>
</tr>
<tr>
<td>Social Media</td>
<td>59% (+11)</td>
<td></td>
</tr>
<tr>
<td>Owned Media</td>
<td>57% (+10)</td>
<td></td>
</tr>
</tbody>
</table>
Search - setting the stage

• 500 000 000 web searches every day

It matters whether search performs well

traditional media
Search

Google - Google Search

bruxelles - query

web results include:
- Belgian Beer Café Bruxelles - Breda
- Brussels - Wikipedia, the free encyclopedia - Brussel
- Cafe Bruxelles - Haarlem
- Cafe Bruxelles: Home
Search
Search is: ranking documents given a query.
Information Retrieval

• Improving …
Information Retrieval

• Improving …
  • query understanding
Information Retrieval

- Improving ...
  - query understanding
  - rankings
Information Retrieval

• Improving …
  • query understanding
  • rankings
  • design
Information Retrieval

• Improving …
  • query understanding
  • rankings
  • design
  • user interactions
Information Retrieval

• Improving …
  • query understanding
  • rankings
  • design
  • user interactions
  • …
Information Retrieval

• Improving …
  • query understanding
  • rankings
  • design
  • user interactions
  • …
• Validation through experimentation
Information Retrieval

What about Observational Studies?

Why not compare with historical data?
Here’s an example of Kindle Sales over time.
You changed the site, and there was an amazing spike

Amazon Kindle Sales

Information Retrieval

External Events can Dwarf Your Changes

- Oprah calls Kindle "her new favorite thing"

- In this example of an A/B test, you’d be better off with version A
- In controlled experiments, both versions are impacted the same way by external events

Amazon Kindle Sales

Information Retrieval

Scaling Experiments at Bing

- We now run over 250 concurrent experiments at Bing

![Graph showing experiments per day]

- We used to lockdown for Dec holidays. No more

Running Controlled Experiments at Scale (1)

Numbers below are approximate to give sense of scale

- In a visit, you’re in about 15 experiments
  - There is no single Bing.
    - There are 30B variants \( (5^{15}) \)
  - 90% of users are in experiments.
    - 10% are kept as holdout

- Sensitivity: we need to detect small effects
  - 0.1% change in the revenue/user metric > $1M/year
  - Not uncommon to see unintended revenue impact of +/-1% (> $10M)
  - Sessions/UU, a key component of our OEC, is hard to move, so we’re looking for small effects
  - Important experiments run on 10-20% of users

Information Retrieval

Running Controlled Experiments at Scale (1)

Search engine research requires experimentation

Evaluation of Search
Evaluation of Search

Which ranking is better?
Evaluation of Search

• Offline
  • Documents annotated for relevance
Evaluation of Search
Evaluation of Search

Annotations are unreliable, expensive
Evaluation of Search

• Offline
  • Documents annotated for relevance
Evaluation of Search

• Offline
  • Documents annotated for relevance

• Online
  • User interactions
Evaluation of Search
Evaluation of Search
Evaluation of Search
Evaluation of Search

Clicks reveal what users prefer
Evaluation of Search

Researchers don't have clicks (yet!)
search the lab perspectives
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks

Diagram:
- Researchers
- Living Labs
- Queries
- Documents
- Ranking
- Clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs for Search
or giving researchers clicks
Living Labs - search engines

- Provider
- Data
- Site traffic
- Info needs
# Living Labs - search engines

<table>
<thead>
<tr>
<th>Provider</th>
<th>Web search</th>
</tr>
</thead>
<tbody>
<tr>
<td>selecion.cz</td>
<td>pre-computed</td>
</tr>
<tr>
<td></td>
<td>document-query</td>
</tr>
<tr>
<td></td>
<td>features</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site traffic</th>
<th>high</th>
</tr>
</thead>
</table>

| Info needs     | vary             |
# Living Labs - search engines

<table>
<thead>
<tr>
<th></th>
<th>Web search</th>
<th>Product search</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Provider</strong></td>
<td>seznam.cz</td>
<td>regiojatek.hu</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>pre-computed</td>
<td>raw queries and</td>
</tr>
<tr>
<td></td>
<td>document-query</td>
<td>(highly structured)</td>
</tr>
<tr>
<td></td>
<td>features</td>
<td>documents</td>
</tr>
<tr>
<td><strong>Site traffic</strong></td>
<td>high</td>
<td>relatively low (~4K sessions/day)</td>
</tr>
<tr>
<td><strong>Info needs</strong></td>
<td>vary</td>
<td>(mostly) transactional</td>
</tr>
</tbody>
</table>
# Living Labs - search engines

<table>
<thead>
<tr>
<th>Provider</th>
<th>Web search</th>
<th>Product search</th>
<th>Local domain search</th>
</tr>
</thead>
<tbody>
<tr>
<td>seznam.cz</td>
<td>regiojatek.hu</td>
<td>uva.nl</td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>pre-computed</td>
<td>raw queries and (highly structured) documents</td>
<td>raw queries and (generally textual) documents</td>
</tr>
<tr>
<td></td>
<td>document-query</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site traffic</td>
<td>high</td>
<td>relatively low (~4K sessions/day)</td>
<td>relatively low</td>
</tr>
<tr>
<td>Info needs</td>
<td>vary</td>
<td>(mostly) transactional</td>
<td>(mostly) navigational</td>
</tr>
</tbody>
</table>
Living Labs - researchers

• 39 teams signed up

• Industry:
  904labs, Microsoft, Plista, Yahoo

• Academia:
  au, bw, cz, fr, ie, in, jp, nl, no, uk, us

• 12 teams submitted rankings
search
the lab
perspectives
Interoperability

• All our efforts are publicly available
Interoperability

• All our efforts are **publicly available**
  
  • code, API’s, plugins, documentation
Interoperability

• All our efforts are **publicly available**
  • code, API’s, plugins, documentation
• We already **collaborate** with another Living Lab (CLEF Newsreel)
Interoperability

- All our efforts are **publicly available**
  - code, API’s, plugins, documentation
- We already **collaborate** with another Living Lab (CLEF Newsreel)
- We are very eager to **work with others**
Interoperability

• All our efforts are **publicly available**
  • code, API’s, plugins, documentation
• We already **collaborate** with another Living Lab (CLEF Newsreel)
• We are very eager to **work with others**
  • *Do you “own” a search engine?*
Cross Discipline
Cross Discipline

- **Currently** Information Retrieval Researchers and computer scientists
Cross Discipline

• **Currently** Information Retrieval Researchers and computer scientists

• Potentially interesting for

  • Behavioral Sciences, Human computer interaction, Design, Information Law, Ethics, ...

  • You?
Transferable Know-how

- Can we experiment with users?
Transferable Know-how

Facebook reveals news feed experiment to control emotions

Protests over secret study involving 689,000 users in which friends' postings were moved to influence moods

Poll: Facebook's secret mood experiment: have you lost trust in the social network?
Transferable Know-how

- Can we **experiment with users**?

- Do **users need to know** that they are part off an experiment?
Transferable Know-how

• Can we experiment with users?

• Do users need to know that they are part of an experiment?

• Do users need to know the aim of an experiment?
Transferable Know-how

• Can we experiment with users?

• Do users need to know that they are part of an experiment?

• Do users need to know the aim of an experiment?

• What data can we collect?
Transferable Know-how

• Can we experiment with users?

• Do users need to know that they are part of an experiment?

• Do users need to know the aim of an experiment?

• What data can we collect?

• Do (public) universities have different responsibility than (private) companies in this regard?
Transferable Know-how

- Can we experiment with users?
- Do users need to know that they are part of an experiment?
- Do users need to know the aim of an experiment?
- What data can we collect?
- Do (public) universities have different responsibility than (private) companies in this regard?
- Can we reuse/share/store experimental data?
Thank You