Simple Techniques for Effective Web Search

Milad Shokouhi
Microsoft Research Cambridge
Web Search Evaluation Based on the User Intent

Milad Shokouhi
Microsoft Research Cambridge
Information Retrieval Evaluation

U.S. oil industry history

Description:
Describe the history of the U.S. oil industry

Relevant documents will include those on historical exploration drilling as well as the history of regulatory bodies. Relevant areas of the oil industry in various states, even if drilling began or later.

Query Results:

<table>
<thead>
<tr>
<th>QID</th>
<th>Title</th>
<th>Exp</th>
<th>MAP</th>
<th>P@50</th>
<th>P@100</th>
</tr>
</thead>
<tbody>
<tr>
<td>701</td>
<td>GX007-22-12200351</td>
<td>2.3091</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX005-75-16559054</td>
<td>1.2432</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX015-23-4395869</td>
<td>0.88791</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX007-11-8691880</td>
<td>0.710262</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX013-49-5273945</td>
<td>0.603673</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX170-56-13957909</td>
<td>0.59224</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX140-74-12844034</td>
<td>0.592206</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX043-76-11448178</td>
<td>0.591642</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX077-21-7464452</td>
<td>0.591167</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX075-65-7127163</td>
<td>0.590749</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX170-46-58399212</td>
<td>0.590371</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX170-80-12023233</td>
<td>0.590323</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX092-16-9574782</td>
<td>0.509699</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX021-06-15644385</td>
<td>0.589195</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX075-62-7095209</td>
<td>0.589106</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>701</td>
<td>GX059-96-18791073</td>
<td>0.588634</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What are you looking for?

Top local listings for pizza hut near Jordan, NY
1. Pizza Hut - Website - (315) 255-2756
   374 Grant Avenue Rd, Auburn - Directions
2. Pizza Hut - Website - (315) 488-1377
   3709 W Genesee St, Syracuse - Directions
3. Pizza Hut - Website - (315) 636-0695
   2265 Downer Street Rd, Baldwinsville - Directions

Local business results for plumbers near Jordan, TX
A. C & C Plumbing - maps.google.com - (512) 303-4101 - more
B. C & C Plumbing - maps.google.com - (512) 303-4101 - more
C. Moncure Plumbing Co - maps.google.com - (512) 303-7586 - more
D. Osborn Plumbing - maps.google.com - (512) 321-6098 - 1 review
E. Parobek Plumbing - www.parobekplumbing.com - (512) 303-4502 - 1 review
F. Heirn Camp Plumbing - maps.google.com - (512) 303-4959 - more
G. Miles & Miles - maps.google.com - (512) 360-3012 - more
H. Potty Doctor - maps.google.com - (512) 466-3938 - more
I. L & R Plumbing - maps.google.com - (512) 303-6410 - more
J. Willis Plumbing - maps.google.com - (512) 303-9273 - more

More results near Jordan, TX »
Which ranking is better?
How about this?
<table>
<thead>
<tr>
<th>Search Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lamborghini</td>
<td>car manufacturer from Italy</td>
</tr>
<tr>
<td>ferrari</td>
<td>Italian luxury car manufacturer</td>
</tr>
<tr>
<td>ferrari usa</td>
<td>website for official ferrari information</td>
</tr>
<tr>
<td>ferrari enzo</td>
<td>specific model of ferrari car</td>
</tr>
<tr>
<td>ferrari pictures</td>
<td>images of ferrari cars</td>
</tr>
<tr>
<td>ferrari pics</td>
<td>images of ferrari cars in specific size</td>
</tr>
<tr>
<td>porche</td>
<td>German luxury car manufacturer</td>
</tr>
<tr>
<td>ferrari images large</td>
<td>large images of ferrari cars</td>
</tr>
<tr>
<td>ferrari wallpaper</td>
<td>wallpapers featuring ferrari cars</td>
</tr>
</tbody>
</table>

**How to decide?**

**Ferrari** - Wikipedia, the free encyclopedia

Ferrari S.p.A. is an Italian sports car manufacturer based in Maranello, Italy. Founded by Enzo Ferrari in 1929 as Scuderia Ferrari, the company sponsored drivers and manufactured race cars before moving into production of street legal vehicles in 1947 as Ferrari S.p.A. Throughout its history, the company has been noted for its continued participation in racing, especially in Formula One, where it has largely enjoyed great success, especially during the 1950s, 1960s, 1970s, late 1980s, and 2000s.

History of Ferrari - The "Cavallino Rampante" - Rosso Corsa - List of models

en.wikipedia.org/wiki/Ferrari - Cached page
WoW!

WOW! Internet Cable & Phone: Home Page
Enter your zip code and click Submit to see if WOW! is available in your area.
www1.wowway.com - Cached page - More info

<table>
<thead>
<tr>
<th>clicks impressions</th>
<th>url</th>
<th>Clickboost score (sort key)</th>
<th>clicks / impressions</th>
</tr>
</thead>
</table>

Blizzard - World of Warcraft
The official site containing news, trailers, gameplay videos, wallpapers, screen shots, and the official forums.
www.worldofwarcraft.com - Cached page
World Of Warcraft - test.worldofwarcraft.com/ptr/charac...
Realm Forums - forums.worldofwarcraft.com/child-fo...
Test Realm Forum - forums.worldofwarcraft.com/board.ht...
Retrieve Password - www.worldofwarcraft.com/loginsupport...
Show more results from www.worldofwarcraft.com
Temporal intent

News results for turkey czech

Turkey - Czech Republic - 17 hours ago
TURKEY 3 CZECH REPUBLIC 2. Turkey go through to the quarter-finals as Group A runners-up and will face Group B winners Croatia in Vienna on Friday 30 + 2 ... guardian.co.uk - 736 related articles
Turkey beat Czech Republic to make Euro 2008 quarter-finals - The Age - 1426 related articles
Turkey, Czech in must-win game, Swiss out of hope against Portugal - Xinhua - 739 related articles

NBC Sports Scoreboard: Euro 2008
matchcentre.stats.com/nbc-euro2008.html - 3k - Cached - Similar pages - Note this

Turkey, Czech in must-win game, Swiss out of hope against ...
BASEL, June 14 (Xinhua) -- Turkey and the Czech Republic must take victory over each other to avoid Euro 2008 elimination in Sunday's match in Geneva, ...
news.xinhuanet.com/english/2008-06/14/content_8367995.htm - 20k - Cached - Similar pages - Note this

Match Centre / Turkey - Czech Republic (15 Jun, 2008)
Get your own news scout! Login and create news tracker to follow your favorite teams and players.
www.piecle.com/livescore/matchreport/turkeyczechrepublic150608/ml4yc3.html - 64k - Cached - Similar pages - Note this

Turkey vs Czech Republic Win Market Betting Odds

Holiday apartments and villas in Turkey
Holiday home rentals, Turkey. Self catering villas and apartments in Turkey - holiday rental accommodation in Turkey including villa rental in Kalkan www.ownerdirect.co.uk/turkey.htm - Cached page

Turkey vs Czech Republic Tickets
Turkey vs Czech Republic Tickets... Buy tickets for Turkey vs Czech Republic today on Seatwave. This is the place to find great tickets for Turkey vs Czech Republic and all other... www.seatwave.com/turkey-vs-czech-republic-tickets/season - Cached page
Welcome to the Official Big Brother site on Network Ten! Get all the latest gossip, exclusive video, chat with the housemates, house updates, news, show information and more.

bigbrother.com.au · Cached page

Big Brother 9 (U.S.) - Wikipedia, the free encyclopedia

Big Brother 9 (also known as Big Brother: 'Til Death Do You Part) is the ninth edition of the American reality television series Big Brother, that aired on the CBS television network from June 12, 2008 to April 27, 2008. Julie Chen returned as host, and Allison Grodner and Rich M�the executive producers.

Pre-season · Broadcasts · HouseGuests · Highlights · Big Brother House · Prizes
en.wikipedia.org/wiki/Big_Brother_9_(U.S.) · Cached page
Summary (part 1)

• Current IR evaluation metrics may not be able to capture the user intent.

• Re-judge the queries.

• New metrics is required for evaluating web search tasks such as image answer triggering

• Pseudo-relevance is important for web search

• Educating the relevance assessors (Bailey et al. SIGIR08)
Evaluating the evaluation measures

- MRR
- MAP
- P@5
- Bpref
- Q-measure
- NDCG
- Rprec
- RBP

\[
\text{AP} = \frac{1}{R} \sum_{i=1}^{\left|R\right|} \left( \frac{r_i}{i} \cdot \sum_{j=1}^{i} r_j \right)
\]

\[
\text{BPref-}k = \frac{1}{R} \sum_{i=1}^{\left|R\right|} \left\{ r_i \cdot \left( 1 - \frac{\min(R + k, i - \sum_{j=1}^{i} r_j')}{\min(R + k, N)} \right) \right\}
\]

\[
\text{Q-measure} = \frac{1}{R} \sum_{i=1}^{\left|R\right|} \left\{ r_i \cdot \left( \frac{2 \sum_{j=1}^{i} r_j'}{i + \min(i, R)} \right) \right\}
\]

\[
\text{nDCG} = \frac{\text{DCG}}{\text{DCG}_I}
\]

where

\[
\text{DCG} = \sum_{i=1}^{Z} g_i / \log(i + 1)
\]

\[
\text{RBP} = (1 - p) \cdot \sum_{i=1}^{\left|R\right|} r_i \cdot p^{i-1}.
\]
Shortcomings...

- Duplication
- Diversity (e.g. query “jordan”)
- Recall
- Ignoring the previous logs (click data etc.)
- Ignoring the UI biases (image answer, snippet quality, position bias).
- No sophisticated user modelling
User modelling

Figure 1: Percentage of times an abstract was viewed/clicked depending on the rank of the result.

Joachims et al. (SIGIR05)
User models vs. evaluation metrics

![Graph showing the comparison between actual and predicted probabilities with error metrics for different models.}]
Side by Side Study

- Metrics
  - RBP
  - NDCG
  - Our preliminary user model
- 6 Users
- 144 queries
- Anonymous results from 2 search engines
## Results

<table>
<thead>
<tr>
<th>User</th>
<th>#Pref.</th>
<th>NDCG</th>
<th>RBP</th>
<th>UM</th>
</tr>
</thead>
<tbody>
<tr>
<td>User1</td>
<td>82/144</td>
<td>0.46</td>
<td>0.48</td>
<td>0.54</td>
</tr>
<tr>
<td>User2</td>
<td>80/144</td>
<td>0.47</td>
<td>0.48</td>
<td>0.57</td>
</tr>
<tr>
<td>User3</td>
<td>65/144</td>
<td>0.55</td>
<td>0.50</td>
<td>0.40</td>
</tr>
<tr>
<td>User4</td>
<td>59/144</td>
<td>0.49</td>
<td>0.54</td>
<td>0.47</td>
</tr>
<tr>
<td>User5</td>
<td>57/144</td>
<td>0.38</td>
<td>0.43</td>
<td>0.61</td>
</tr>
<tr>
<td>User6</td>
<td>56/144</td>
<td>0.48</td>
<td>0.62</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td><strong>399</strong></td>
<td><strong>0.47</strong></td>
<td><strong>0.51</strong></td>
<td><strong>0.52</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.21</td>
<td>0.61</td>
</tr>
</tbody>
</table>
Summary (Part 2 -- final)

• Hard to make solid conclusions about metrics
• NDCG does not look particularly good in the side by side study.

Future work:

• Adding more queries
• Adding more judges
• Adding more features to the user model