Toward higher quality health search results:
Automated quality rating of depression websites

David Hawking
CSIRO ICT Centre

Medinfo Workshop - Brisbane 19 Aug 2007
A Multi-way Collaboration

- ANU Centre for Mental Health Research
  - Kathy Griffiths & Helen Christensen
- CSIRO ICT Centre
  - David Hawking & Peter Bailey
- ANU Computer Science Department
  - Tim Tang (PhD thesis)
  - Ramesh Sankaranarayana
- Microsoft Research
  - Nick Craswell

**depression → obesity**
Why Study Depression?

- Depression is the leading cause of disability burden in Australia. *(AIHW)*
- One in five people suffer from a mental disorder in any one year. *(ABS)*
- Approximately one in five suffer clinical depression in their lifetime.
- Depression is the leading risk factor for suicide.

But ... 

- Only one in three people with a mental disorder receive treatment.
- Many who do receive treatment do not receive evidence based treatments.
The AQA procedure

Described in Griffiths et al, JMIR 7(5), 2005

- Start with evidence-based site ratings
  - Guidelines from Oxford Centre for Evidence Based Mental Health
  - Systematic ratings - scale of 0-20
  - 29 sites
- What words and phrases tend to distinguish high quality documents from low quality ones?
Learning a “Quality” Query

HQ
110 documents judged relevant and of high quality

Other
3002 documents, either irrelevant or relevant but low quality

<table>
<thead>
<tr>
<th>Term</th>
<th>HQ</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>depression</td>
<td>10.3</td>
<td>ECT</td>
</tr>
<tr>
<td>treatment</td>
<td>5.7</td>
<td>antidepressants</td>
</tr>
<tr>
<td>disorder</td>
<td>3.3</td>
<td>zoloft</td>
</tr>
<tr>
<td>patient</td>
<td>3.3</td>
<td>mental health</td>
</tr>
<tr>
<td>medication</td>
<td>3.0</td>
<td>cognitive therapy</td>
</tr>
</tbody>
</table>
Learning a “Relevance” Query

- As for Quality Query, but using 347 documents relevant to the general topic of depression and 9,000 with very low probability of relevance.
1. The websites are crawled.
2. The crawled pages are aggregated with a collection of arbitrary ordinary documents and indexed.
3. The relevance query (described above) is run against the index and, for each candidate site, the number $|R|$ of retrieved documents and the mean relevance score $\bar{r}$ are calculated.
4. $|Q|$ and $\bar{q}$ are analogously calculated for the quality query.
5. Site relevance and site quality scores are calculated using Equations 1 and 2.
6. An overall site score in the desired range of 0-20 is calculated using Equation 3
Mathematical details

\[ S_r = \alpha \bar{r} + (1 - \alpha)|R| \]  \hspace{1cm} (1)
\[ S_q = \alpha \bar{q} + (1 - \alpha)|Q| \]  \hspace{1cm} (2)
\[ S = \gamma(\beta S_q + (1 - \beta)S_r) \]  \hspace{1cm} (3)
AQA prediction of expert ratings

Mean evidence-based score vs. AQA score

- Observed data
- Linear fit
- Quadratic fit

Testing on 29 human-rated sites not used for training.
But couldn’t we just use Google PageRank?

Testing on 29 human-rated sites not used for training
AQA scores are pretty good

How can we use them?
1. Guiding a *quality-focused crawler* to fetch pages for a search portal
2. Filtering or reranking Google/Yahoo!/Live results for depression-oriented queries.
3. A rating service for consumers.

We’ve confirmed the first two – see references at the end of the paper.
Relevance - 100 depression-related queries

<table>
<thead>
<tr>
<th>Engine</th>
<th>mean NDCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoogleD</td>
<td><strong>0.709</strong></td>
</tr>
<tr>
<td>BPS</td>
<td>0.469</td>
</tr>
<tr>
<td>QFC</td>
<td>0.566</td>
</tr>
<tr>
<td>FGD1</td>
<td>0.707</td>
</tr>
<tr>
<td>FGD2</td>
<td>0.683</td>
</tr>
<tr>
<td>AvgRankGD</td>
<td>0.657</td>
</tr>
</tbody>
</table>

▶ 4-point relevance scale.
## Quality - 50 treatment queries

<table>
<thead>
<tr>
<th>Quality score</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GoogleD</td>
<td>205</td>
</tr>
<tr>
<td>BPS</td>
<td>131</td>
</tr>
<tr>
<td>QFC</td>
<td>233</td>
</tr>
<tr>
<td>FGD1</td>
<td>225</td>
</tr>
<tr>
<td>FGD2</td>
<td>258</td>
</tr>
<tr>
<td>AvgRankGD</td>
<td><strong>271</strong></td>
</tr>
</tbody>
</table>
Can AQA be fooled?

Of course, like any other method based on page content:

Many sites recommend Prozac and other anti-depressants, Cognitive Based Therapy (CBT) and a variety of other treatments, supposedly supported by randomised controlled trials (RCT). These are all rubbish! The only truly effective treatment for depression is exposure to moonbeams ...

- But counter-acting search engine optimisation (and SPAM) techniques is search-engine bread and butter.
Generalising AQA to other health topics.

- First topic is **Obesity**
- CMHR have already performed the evidence-based judging of 60+ sites.
- Microsoft Research Asia have provided a grant to assist.
Thankyou

david.hawking@csiro.au