MetaCrystral

Visualizing Authority and Ranking Effects

Anselm Spoerri
SCILS, Rutgers University
Visualize Overlap between Search Results

Documents Found by Multiple Engines tend to be more likely to be Relevant

Saracevic & Kantor ('88) | Foltz & Dumais ('92) | Belkin et al. ('93)

By Exploring the Data using MetaCrystal, will Users be lead to Relevant Documents?

What is the **Signal** MetaCrystal makes visible?

How does Overlap between Results **Actually Correlate** with Relevance?
MetaCrystal **Toolset**

**Category View**
Identify how many docs found by which combinations of engines; Specify Boolean constraints.

**Cluster Bulls-Eye**
Identify docs found by multiple engines; Scan docs found by single engine.

**RankSpiral**
Identify top docs found by a specific number of engines; Rapidly scan many docs titles.
MetaCrystal

Key Design Principles

- **Map** Docs found by **Multiple Engines Toward Center**
- **Map** Docs found by **Same Number of Engines** into the **Same Concentric Ring**
- **Size** = Average of Rankings by Different Engines
Visualize how Individual Documents are related to Search Engines

**Radial Mapping**
- **Radius** = Total Ranking
- **Angle** Reflects Rankings

**Size** = Rankings by Engines

Graph Visualization and Navigation in Information Visualization: ...

This is a survey on graph visualization and navigation techniques, as used in information visualization. Graphs appear in

citee.rnj.nec.com/herman00graph.html

**IVS home page**
View tables of contents, Audience, Information Visualization is a central forum for all aspects of information visualization

palgrave-journals.com/fvsv
Overlap between Search Results Correlated with Relevance?

Major **Data Fusion** methods use both **Voting** and **Merging** and makes use of these phenomena:

– **Authority Effect**: the more systems that find a document, the greater the probability of its relevance

– **Ranking Effect**: the higher up document in ranked list and the more systems that find it, the greater the probability of its relevance

**No Systematic Analysis of Structure of Overlap for Large Text Databases** as used in TREC
Overlap between Search Results Correlated with Relevance?

**Method**

- Use Ad-hoc track data for TREC 3, 6, 7, 8
- Systems search the SAME Database
- Automatic Short Runs
- **50 Topics** and **1,000 Documents per topic** → 50,000 documents
- Retrieval systems can submit multiple runs
  - Select Best Run based Mean Average Precision

<table>
<thead>
<tr>
<th>TREC</th>
<th># Systems</th>
<th>Documents Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>TREC 3</td>
<td>19</td>
<td>928,709</td>
</tr>
<tr>
<td>TREC 6</td>
<td>24</td>
<td>1,192,557</td>
</tr>
<tr>
<td>TREC 7</td>
<td>28</td>
<td>1,327,166</td>
</tr>
<tr>
<td>TREC 8</td>
<td>35</td>
<td>1,723,929</td>
</tr>
</tbody>
</table>

- **Compute Average** by summing over all 50 topics and divide by 50
How does Overlap Correlate with Relevance?

To Appear in ASIST 2005

Authority Effect
TREC 8 – Impact of Average \textbf{Rank Position}?

Compute overlap structure between top 50 search results of 35 random groupings of 5 retrieval systems for 50 topics.

\section*{Ranking Effect}

Submitted to InfoVis 2005
How does Overlap between Search Results Correlate with Relevance?

**Authority Effect** – the more systems that find a document, the greater the probability that it is relevant

**Ranking Effect** – the higher up a document in a ranked list and the more systems that find it, the greater the probability of its relevance

- Validates MetaCrystal’s Design Approach
- MetaCrystal Visualizes Authority & Ranking Effects
- MetaCrystal can Guide User’s Exploration Toward Relevant Documents