Measuring the value of search trails in web logs

Presentation by
Maksym Taran & Scott Breyfogle

Research by Ryen White & Jeff Huang
What's all this?

• When we search, we usually get individual documents
• But people often click around to other pages…
• **Search trails!**
• Paper looked at whether they were useful or not
• Got data, analyzed ‘usefulness metrics’
• Ostensibly to make search results pages better
Search trails

Here’s one:

Components: Origin, Destination, Sub-trail
The data

- Data from 3 months of user logs
- 2 months for user history, 1 month for analysis
- Pruned heavy and light users
- Only included queries with relevance ratings
- Took at most 10 trials / person
The experiment

• Computed metrics for the remaining data
• Compared metrics
  – between trail components
  – between popular and unpopular queries
  – user query history
Metrics: Relevance

• Human Rating
• Six point scale
• Average different views
Metrics: Coverage

• Background info
  – DMOZ / Open Directory Project

• Query interest model
  – consists of the DMOZ labels for query results
  – weighted so label frequency counts

• Coverage of a site w.r.t. a query:
  – fraction of weighted labels relevant to the query that are also relevant to the site
Metrics: Diversity

• Diversity of a site w.r.t. a query:
  – Unweighted coverage
  – **Unweighted** fraction of labels that are relevant to a query that are also relevant to the site
Metrics: Novelty

• Novelty of a site w.r.t. a query and a user:
  – Diversity, but taking into account user history
  – The \textbf{unweighted} fraction of relevant labels that a user hasn’t already seen for this query
Metrics: Utility

• Hard to get a good metric for this
• An easy one is time spent on site
• Utility of a site w.r.t. a query:
  – did the user spent 30 s or more on the site?
  – a source said it was reasonable…
Results!
Results! Query popularity

• Queries were split into 3 popularity tiers

• Saw all metrics increase with popularity

• Agrees with previous findings that search engines perform better for more popular queries
Results! Query history

- Some users had histories of re-running queries
- Queries were grouped by ‘historical frequency’
  - None
  - Some
  - Lots
- Relevance and utility rise with more history
- Coverage, diversity, novelty fall
Criticism

• Most results trivial:
  – More topics are covered when more than one page is visited
  – First and last pages are more relevant
    • User knows what they are looking for
• Definition of a session is shaky
  • Many search paths may end after the first few clicks
  • Users recycle tabs
Criticism & Further work

• Questionable applicability to results pages
  – It’s difficult to show trails on results pages
  – Do the users **value** the extra coverage/diversity/novelty in the trail?

• Further work needed
  – Figure out whether it would be good to show people trails
  – Find more appropriate/user-derived metrics