FollowMyLink

Individual APT Presentation First Talk February 2009

Overview

- General Principles
- Reading List
- Tasks involved
- Schedule

General Principles

- Plagiarism
- Referencing
- <u>ACM Digital Library</u>, <u>IEEE Digital Library</u>, <u>Springer-Verlag Digital Library</u>

Reading List

- Kraft, R., *et al.*, "Y!Q: Contextual Search at the Point of Inspiration", in Proceedings of CKIM'05, 2005 <u>http://www.soe.ucsc.edu/~rekraft/papers/yq_cikm.pdf</u>. Also see <u>http://yq.search.yahoo.com/publisher/firefox.html</u>
- Lieberman, H., *et al.* Why Surf Alone? Exploring the Web with Reconnaissance Agents", in Communications of the ACM, August 2001, pp. 69-75.
 <u>http://web.media.mit.edu/~lieber/Lieberary/Letizia/Why-Surf/Why-Surf.doc</u>
- Finkelstien, L., *et al.*, "Placing Search in Context: The Concept Revisited", in Proceedings of WWW10, 2001. <u>http://www.cs.technion.ac.il/~gabr/papers/context_search.pdf</u>
- Mozilla Development Center, 2006, "Building an Extension"., <u>https://developer.mozilla.org/en/Building_an_Extension</u>

Nelson vs. W3C

- Definitions of hypertext
 - 'Well, by "hypertext" I mean *non-sequential writing*--text that branches and allows choices to the reader, best read on an interactive screen' Ted Nelson, 1987. Literary Machines, Edition 87.1.
 - "Hypertext is text which is not constrained to be linear. Hypertext is text which contains links to other texts." <u>http://www.w3.org/WhatIs.html</u>
- Static vs. Dynamic₁ links vs. Dynamic₂ links
- Dynamic₁ is computed end of fixed source anchor
- Dynamic₂ is computed end of user-created source anchor (can we call it an "ephemeral link"?)





Tasks

- User Model
- Identifying User Interests
- Making Recommendations
- User Feedback
- Evaluation

Task: The Page & User Models

- Page Model
 - What is the current page about?
- User Model
 - What is the user interested in at the time that the user selects text to FollowMyLink?
 - Can be extracted from the page model
 - Can be extended over a sequence of pages the user has already visited (and followed links or FollowMyLink links from)

Task: The Page & User Models

- At least an attribute-value pair list.
 - More complex user model representations may be used.
- UM guides the adaptation process.
- Where do the models reside?
 - Either "Outside" of the adaptation process
 - Or part of the browsing environment (e.g., part of a Web browser)

Task: The Page & User Models

- When are the models updated?
 - PM every time a new page is loaded
 - UM every time the user selects text and FollowMyLink
- From where do we get the attributes?
- Should the attribute values be "aged"?
- How do we 'reset' the UM?
- How do we include normally followed links in the UM?
- Should we include links from the bookmarks, keyed in directly?

Task: Identifying User Interests

- How can we distinguish between what a user is interested in, and what a user is not interested in?
 - Or at the very least, what evidence is there that a user is interested in something?
 - The page that the user is visiting. All of it?
 - The anchor text of the link the user clicked on to get here?
 - The region surrounding the link?
 - The heading/title/other descriptive text surrounding the link that the user followed
 - If attribute-value pairs are used, a *term* can be used as an attribute, and its value can either be binary or real.

Task: Identifying User Interests

- Once the user's interests have been identified the user model can be updated.
- Should all terms be used, or only meaningful and significant ones? Word "as is" or root? And how do we submit query to search engine if root?
- Understanding where the user is in relation to his or her browsing pattern (optional).
 - Do we need to know if a user has backed up from a page? If we assumed that the user was interested in the info, and we added it to the user model, do we now need to remove it?

Task: Making Recommendations

- A suitable description of the user's interests must first be derived from the user model
 - For example, pick top *n* scoring terms from the user model if real values are used, and value exceeds some threshold
 - How can one distinguish between important terms if binary values are used?

Task: Making Recommendations

- Terms in the description are combined to form a *query* that can be submitted to a search engine, e.g., Google
- Page Recommender:
 - Submit query to Google and go to most relevant Web page
 - Check that 'best' page is not already current page, or page seen recently!
 - These recommendations should be:
 - User settable the user may choose to turn off this feature
 - If user asks to see all relevant pages, then present them in a separate window (or frame, if preferred). If a user clicks on a recommended page, the page loads in a new window.

Task: User Feedback (optional)

- Documentation must contain section about this, but implementation is optional
- Should users be able to provide feedback about the system's performance?
- If so, how? And how will feedback change the system's performance?
 - That is, how will the system learn from User Feedback?

Evaluation

- Ideally, system is evaluated.
- How can the system be evaluated, to ensure that users are generally satisfied with the pages that they're shown?
- Compare your results to normal Web search and Y!Q search. Can you explain the differences, if there are any?

Pitfalls

- Selecting an image/non-textual data and then selecting "FollowMyLink"
- What should happen if the 'best' page according to Google is missing/cannot be loaded?
- What should happen if Google makes no recommendation?
- Should the user be able to modify the query submitted to Google?

Schedule

- Until w.c. 19th Feb inc: Discussion, talks once/week
- w.c 16th March: Submit Table of Contents/chapter overview for feedback (optional)
- w.c. 24th Apr: Demo 1 (optional)
- 20th Apr-11th May: Submit one chapter of your choice for feedback (optional)
- w.c. 4th May: Demo 2 (optional)
- May 22: Submit APT report
- June: Demo under exam conditions