

# FollowMyLink

Individual APT Presentation

Second Talk

February 2009

# Overview

- Identifying user interests
- Topic segmentation
- Adding terms to the user model (intro)

# 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?

# What user interests can be captured?

- **According to Brusilovsky** (Brusilovsky, P. (1996). *Methods and Techniques of Adaptive Hypermedia*. In User Modelling and User-Adapted Interaction, 6(2-3), pp. 87-127. Available on-line at <http://www.contrib.andrew.cmu.edu/~plb/UMUAI.ps>.)
  - User knowledge
  - User goals
  - User preferences
  - User background/experience

# User Knowledge

- ITS approach
  - User knowledge, expert knowledge/domain knowledge
  - Capture differences (as user model) using e.g., overlaymodel
  - Stereotype UM: need to know enough about user to classify him/her

# User Background/experience

- Usually provided directly or indirectly by user
  - Questionnaire
  - Test
- User's self-assessments are frequently inaccurate

# User preferences

- Can sometimes estimate, but usually supplied by user



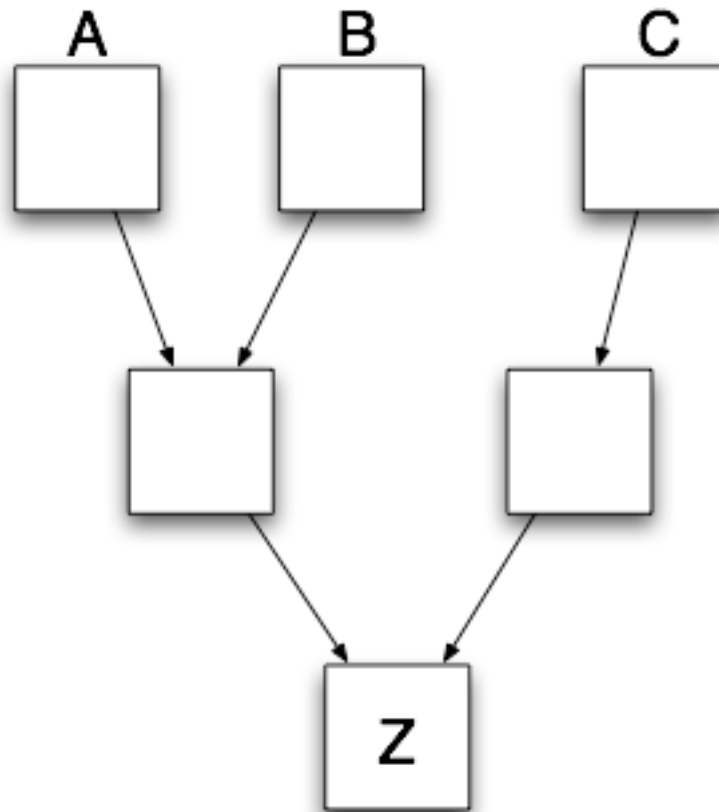
# User goals

- What is the user trying to achieve (e.g., find) and (perhaps) why?
- In FollowMyLink we know only where the user has been recently
- We know nothing about the domain and nothing about the user's knowledge of the domain (although we could also have access to bookmarks/favourites)

# User goals

- Brusilovsky claims that the user's goal is the most changeable user feature: almost certain to change from session to session (and possibly even within the same session)
- Adaptive navigation support generally applied to help users achieve their goals.

# Following different paths



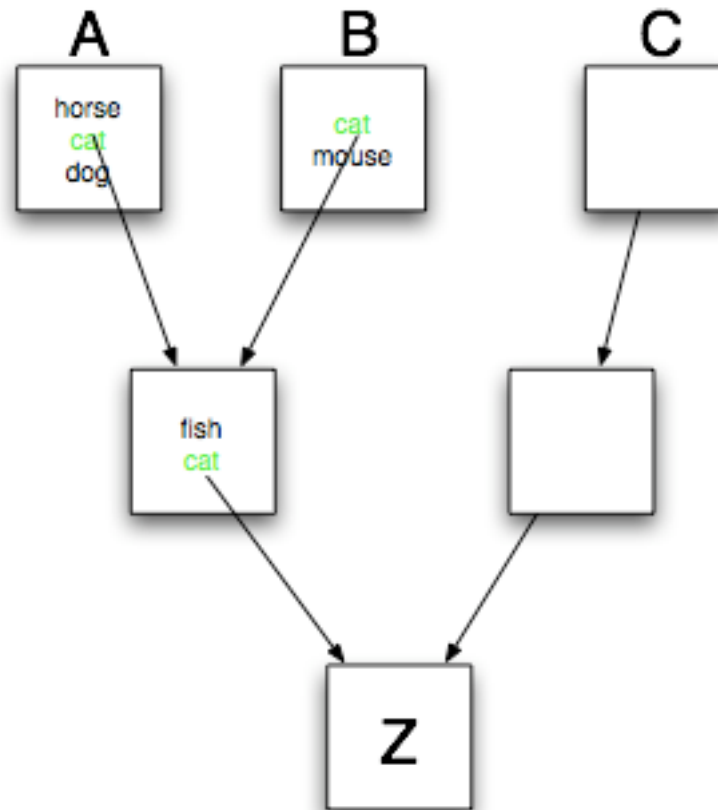
# Different interests?

- Do users A, B, and C necessarily have the same interest in node Z?
- How can we distinguish between the interests of the different users?
- Do we simply take all the terms from the last few pages that the user has encountered, use a window of text around the area the user is interested in, or can we be more accurate?

# Topic segmentation

- A couple of different approaches (see about or look for Freddy Choi's text segmenter, Marti Hearst's TextTiling, and HyperContext)
- Choi's segmenter and TextTiling approaches break up a text into different sections whenever a new topic is detected
- HyperContext creates *interpretations* of a web page based on its parents

# Topics and user interests



# Topics and user interests

- If we simply take all of the terms that occur in each visited page, then we may add terms to the user model that the user is not necessarily interested in
- But how do we know in which part of the page the user is interested?

# Indications of user interest

- Follow a link...
- Query provided to search engine...
- Select text and request a FollowMyLink...
- Any others?
- Extract relevant passage using DOM, for instance
  - Should we add terms in nearby, relevant headers/  
title?



# Adding terms to the User Model

- Once we have the terms in which the user is interested we can add them to the user model
  - which is empty to begin with
- Should we add all the terms unmodified?
  - Or can we eliminate non-content bearing words (i.e. ‘stop words’)?
  - And should we add only word stems (so that morphologically related terms are recognized)?

# Adding terms to the User Model

- How do we identify which (of the remaining) terms are more important so that we can weight them accordingly?
  - Important, because we don't necessarily want to submit very long queries to the search engine
- Should we 'age' terms that have been in the UM for a while, but haven't been recently encountered in the pages visited by the user?

- More next week...