



Evaluating a General-Purpose Adaptive Hypertext System

Chris Staff

Department of Computer Science & AI

University of Malta



Aims of this presentation

- To briefly discuss HyperContext
- To discuss the *process* of evaluation, rather than the results themselves
 - Intended process
 - Actual process
 - Lessons learned



Aims of HyperContext

- General-Purpose Adaptive Hypertext System
- To provide a “cheap” but reasonably accurate method of achieving adaptivity
 - In a domain independent environment
- To achieve adaptivity as a user browses through hyperspace
 - To provide support for short-term interests



General-purpose AHS

- Unlike ITS-based AHSs, we do not have:
 - Known user goals
 - “Check-list” of what user knows compared to what user doesn’t know, or wants to know
 - Deep domain model
 - Homogeneous information



General-purpose AHS

- We discover user interests
 - As a user browses
 - Without prior knowledge of a user's interests
 - With only a surface representation of the document collection



Premise

- Adaptivity is required because the same document may mean different things to different people
 - The same document may contain information on different topics, which may be relevant in different situations
 - The same piece of information may even convey different, and possibly mutually exclusive, meanings to different people



The screenshot shows a web browser with three windows:

- Hypertext Books:** Contains text about a book of pre-Web hypermedia designs and a link to "From Memex to Hypertext : Vannevar Bush and the Mind's Machine".
- Vannevar Bush:** Contains a photo of a desk with a microfilm storage device and text describing the Rapid Selector and Memex machine.
- As We May Think:** Contains the title "As We May Think" by Vannevar Bush, a portrait of Vannevar Bush, and a paragraph about the article's original publication in The Atlantic Monthly.

Red and blue arrows indicate hyperlinks from the text in the 'Hypertext Books' and 'Vannevar Bush' windows to the 'As We May Think' window.



Why are links created?

- Usually because the author believes the destination document contains information useful to the reader
- Presume that all readers of the same document require the same information from it, or that they will “help themselves” to information chunks as appropriate



Interpretations

- HyperContext facilitates the creation and re-use of explicit *interpretations* of information



Interpretations

- Interpretations are partial representations of the information a document contains
- An interpretation is represented as a vector of term-weights
 - Term weights may differ in each interpretation
 - Allows query-based retrieval of interpretations using a third-party IR system (e.g., SWISH-E, Google)



Partitioning Hyperspace

- We also associate links and link destinations with interpretations
- Links relevant to one interpretation may not be relevant to another interpretation of the same document
- Likewise for the destinations of links
- Explicit interpretations allow us to build a model of a user's short-term interests



Summary

- HyperContext hyperspace constructed by connecting explicit *interpretations*
- Interpretations are created by users who create links
- We assume that a link destination is relevant to a source document
 - Implicit in “normal” hypertexts, but a stronger assumption in HyperContext
 - Source of the “context” in HyperContext



Evaluation Process: Objectives

- That relevant information can be found “faster” in HCT hyperspace than in a “without-adaptivity” equivalent
- That relevant documents can be recommended based on a short path of traversal through hyperspace



Evaluation Process: Issues

- Uncontrolled vs. Controlled hyperspace
 - More robust prototype required for uncontrolled version
 - Harder to motivate participants in controlled version
 - Associated costs
- We converted part of www.w3.org
 - Likely to be free from HTML error
 - Consistent



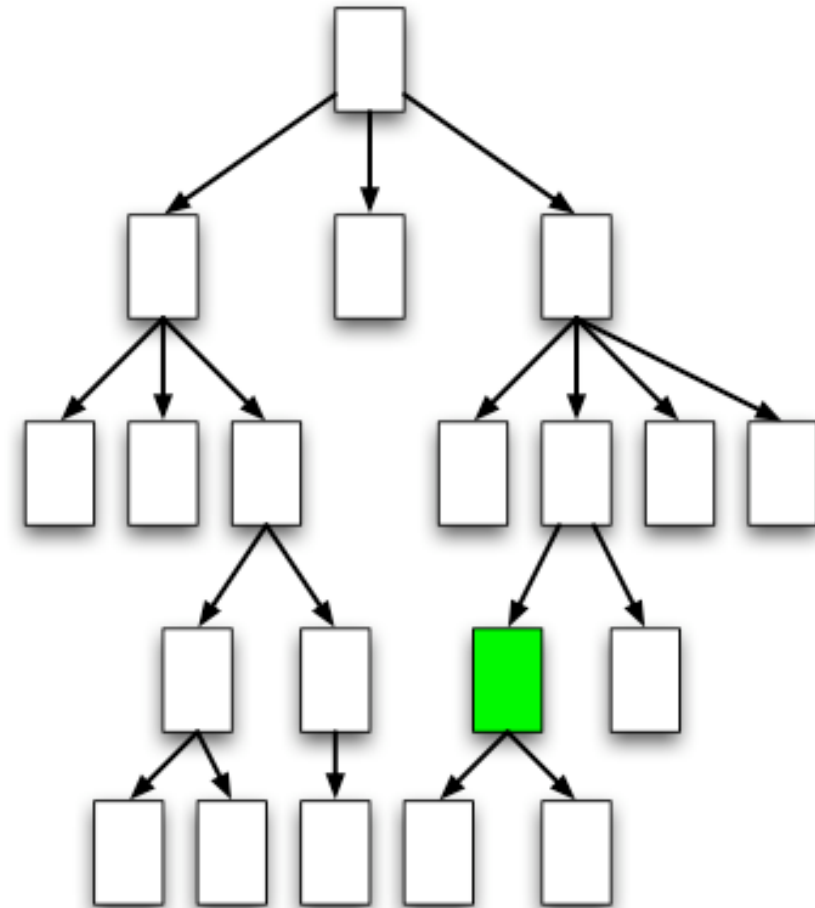
Evaluation Strategy

- Wanted to involve human participants to assess speed with which relevant information can be found
- Settled upon two mostly automated strategies
 - How “quickly” can some randomly selected document be located using link ordering algo?
 - Relevance of HCT recommended document to others in path of traversal



Locating documents

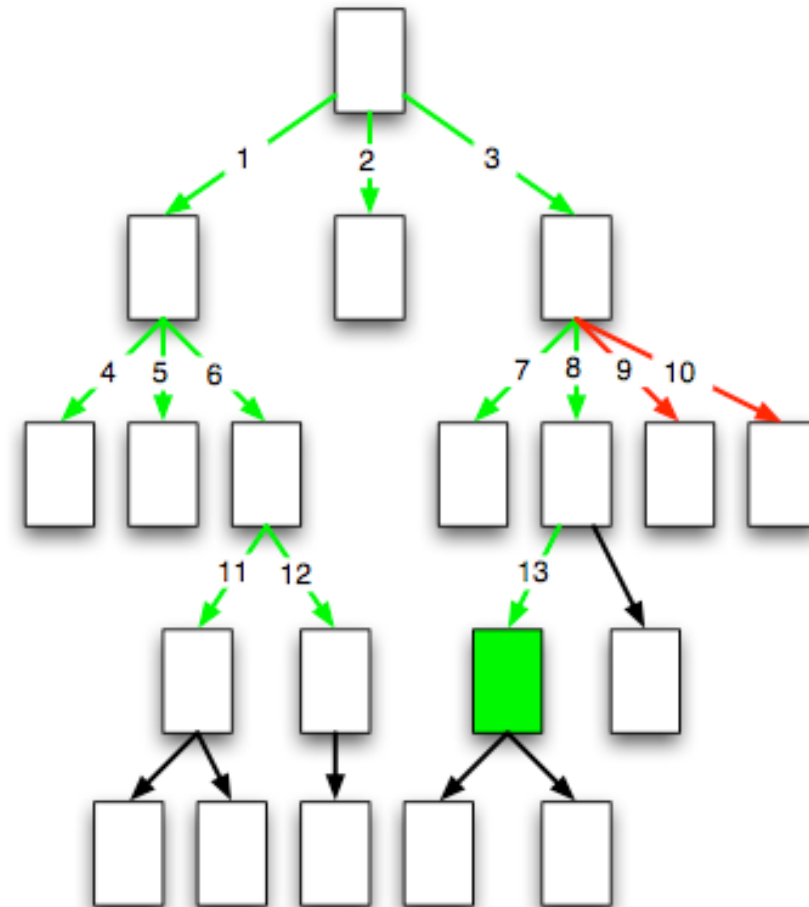
- Select a target document at random





Locating documents

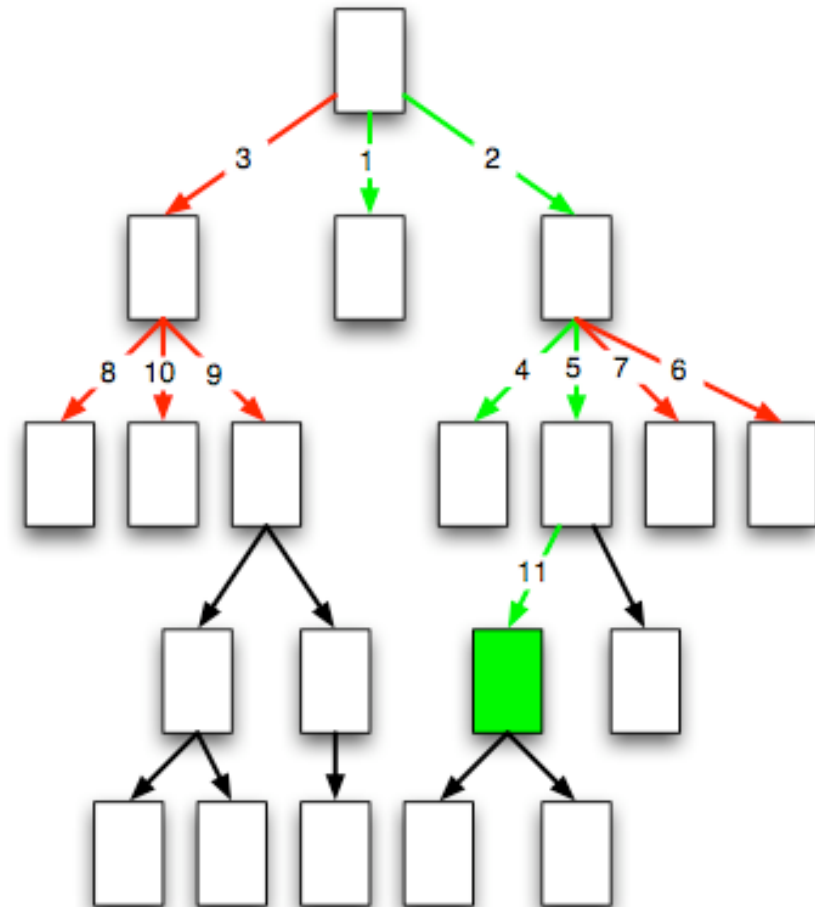
- Starting from root document, follow links using breadth-first approach vs. link ordering algo
- Compare interpretation of destination to target doc (using cosine similarity)





Locating documents

- Optimal solution would choose shortest path, worst would choose longest path





Commentary on results

- Better results when the target document is further from the root document (3 or 4 link traversals)



Document Recommendations

- 11 randomly selected paths of 4 link traversals
 - max path length reported from previous evaluation
- Automatically constructed two user models as if user had travelled the path
 - adaptive: used HCT document interpretations
 - control: used full document



Document Recommendations

- Derived query from each user model, submitted it to IR system
 - IR system contains index of interpretations and full docs of converted www.w3.org site
- Selected top ranked document from each and “packaged it” (in random order) with path



Document Recommendations

- Placed each of the 11 paths + 2 doc recs per path on-line
- Invited students and staff of Dept. CS&AI, UoM to give relevance judgements on recommended documents
- Used 4-levels of subjective relevance
 - Highly R, quite R, quite NR, highly NR



Commentary on results

- At face value, documents recommended by control were considered more relevant
- However, time spent on a path was recorded
- *If* this is indicative of reading time, *and* this is indicative of skim- or deep-reading then adaptive rec preferred by deep-readers, and control rec preferred by skim-readers



Lessons learned

- Experimental set-up is expensive if “real” participants are used
 - Either robust system needed
 - or need to recruit *really* interested participants
- Semi-automating experiments can be beneficial



Conclusions

- We would have preferred to use “real” participants
- Ideally, test collection with relevance judgements is available (*a la* TREC)
 - Especially useful to evaluate user models that do not use/collect info from outside system...
 - ... or when stereotypes are used



Thank you!