

## PROJECT ORGANIZATION

Our class project will be **Knowledge-based Hyperlinks (KBHL)**. (The name is suggestive, not final.) The general idea is to develop a demonstration example (simulation of) traversing a network of hyperlinks based on content or semantic information, rather than on syntactic structures such as keywords.

The component tasks/roles for this project include:

0. Research: have others tried this approach? What did they learn? What topics partially address our project (eg: knowledge engineering, hypertext, web search, expert systems, interactive interface, software agents, etc.)?
1. Develop a sample database of content-carrying web-pages, with hyperlinks across various content components.
2. Develop a knowledge-based data-structure which attaches to each hyperlink. The knowledge-base will contain conventional expert system-like assertions of facts and relations.
3. Develop an inference strategy for traversing the knowledge of various hyperlinks. This may include pattern-matching, Bayesian probabilities, various types of inference, and other semantic-like structures. See below for more ideas.
4. Develop an interaction plan which permits the user to understand and traverse knowledge-based links.
5. Develop an interface prototype for using KBHLs. The interface should require minimal learning, and have a “natural” feel.
6. Discuss and roughly design extensions to the KBHL, including user-extensibility, automated documentation, and software agency.

## Discussion

The content web-pages require careful selection and design, to maximize the (apparent) utility of the KBHL tool. Research from Ontology Engineering (ie what the folks at Yahoo do to organize their weblinks) will guide this effort.

The available types of *intelligent traversal* require careful design, so that 1) the tool will be useful for finding information, and 2) the tool will be understandable to normal folks.

How semantics is captured and accessed is of critical importance. How do we know what the user is looking for? How will users be able to say what they are looking for? What types of intelligent traversal are useful?

A given links will usually contain may intelligent branches. How will the user know which to select? That is, the organization of information is not only, or even necessarily, logical.

## Programming the Interface

What type(s) of organizational structure do we want our smart links to expose? Inferential techniques address implicit or embedded information. Other types of smart links expose different structures. For example,

a *refinement link* leads to more detail on a topic.

a *classification link* provides a property inheritance context.

a *chronological link* tells you what happened before or after.

a *spatial link* navigates through locations.

a *dependency link* identifies prerequisites, requirements and causal structures.

a *structure link* decomposes an object into its component parts.

a *decision link* traces choices and their consequences.

an *analogy link* identifies things that are similar but not necessarily related.

Our problem maps onto a classic graph problem: what kind of nodes and vertices make sense? How many types of links can be used at one time? Should nodes or links provide consistency?

### Observations

The simulation web-pages need to accurately reflect the data-structures underlying actual web-pages. We will need to figure out how additional information can be easily and portably attached to links.

Specialized types of inference are needed for different fields of knowledge. Only some kinds of knowledge are reducible to knowledge-based encoding.

Knowledge may not be about “content”, it can also be about structure (the form of the link), about possibilities, about grouping, about proximity, etc.

We will need to show *critical functionality*. What does our tool do that other tools do not do? How is the advantage measured? Where are the strong, weak, and failure points?

There may be no solution for information overload. We can be overwhelmed by too many windows, by too many nodes and links, by too much scrolling. Perhaps links should filter and refine rather than enhance access.

Understandable structure may need to be designed and written into the website itself, rather than put into links.

Techniques for structuring and filtering:

*Labeling*: clear, concise labels and concepts

*Chunking*: relatively small, related hunks of information

*Relevance*: all information pertains to the content of the page or the goal of the user

*Consistency*: similar items are treated in similar ways

Hyperlinks may simply increase the desire for better content structure and more efficient linking models. That is, smart links may expose the greater weaknesses of hypertext systems.

Bottom line is that the information itself must have a structure for a smart link to expose.