Lin, X., & Chan, L. (1997). Knowledge Class' - A dynamic structure for subject access on the web. 8th ASIS SIG/CR Classification Research Workshop, 31-40. doi:10.7152/acro.v8i1.12726

Knowledge Class -- A dynamic structure for subject access on the web

Xia Lin
College of Information Science and Technology
Drexel University
Philadelphia, PA 19104-2875

Lois Mai Chan
School of Library and Information Science
University of Kentucky
Lexington, KY 40506-0039

Abstract

This project investigates how traditional information organizing concepts, methods, and tools may be applied to the digital environment, both theoretically and practically. A new method of knowledge organization in the distributed environment is developed. The method, Knowledge Class, is designed to be both an information organizing device and an information access tool. Individually, each knowledge class is an organized entity dynamically linked to web information resources. Collectively, instances of knowledge classes can be assembled through classification hierarchies to form a conceptual infrastructure for organizing and managing web resources. The framework for Knowledge Class is a Java/JavaScript interface that integrates recent web technology with traditional devices of knowledge organization and access.

Introduction

Although the richness of electronic resources available on the World Wide Web is gratifying, the web's size and range are often overwhelming to those searching for information. There are currently many tools that enable users to find "some" information on a particular topic. However, for many users, seeking specific information on the web and maintaining it for future use can be daunting and time-consuming. Achieving precision, in particular, poses a real challenge.

Because information in this new digital environment is distributed, there is nothing analogous to the access and location provisions of a centralized database. As a result, the essence of the digital environment is no longer the information it stores or possesses, but the richness of the links it provides. At the early stage of World Wide Web, bookmarking was a popular approach for the user to keep track of interesting sites (Notess, 1995). As the web has grown, there have been many efforts devoted to developing organizing schemes beyond bookmarks, including the classification approach (McKiernan, 1997), and the Metadata approach initiated by OCLC Online Computer Library Center (Weibel, 1995). Significant developments in robot indexing (Koster, 1995) also led to commercialization of many search engines such as Yahoo!, AltaVista, and Lycos, InfoSeek, etc. Like traditional information organizing tools such as classification systems and thesauri, these search engines and organizing schemes have been designed for all

ISSN: 2324-9773

Lin, X., & Chan, L. (1997). Knowledge Class' - A dynamic structure for subject access on the web. 8th ASIS SIG/CR Classification Research Workshop, 31-40. doi:10.7152/acro.v8i1.12726

users. They attempt to organize the whole World Wide Web resources. They are extremely useful for identifying some information on general areas; but those who needs specific information in well-defined, finite domains. What is needed are flexible and adaptable mechanisms by which individual users can create and maintain personalized maps of promising resources by organizing and managing the links that lead to them. These maps can be easily adapted to evolving information needs and to changes in the digital environment. Furthermore, for those who are not familiar with sophisticated retrieval methods or who lack sufficient computer expertise, there is also a need for improved means of both accessing and navigating remote electronic resources.

Purpose of Project

The purpose of our project is to explore a new method for customizing knowledge organization and access on the World Wide Web. It proposes and tests a new device for organizing and delivering information resources on the web for individual users. This device consists of an information organizing framework with a web interface, designed to respond to individual research or personal interests, with dynamic as well as static links to the web. When fully developed, it is hoped that the information organizing framework and the interface will help individual researchers and serious information seekers to retrieve, organize, "store," and manage electronic resources for individual use.

We propose a concept called "Knowledge Class,"[8] which serves as a conceptual building block for web access. It is similar to a classified mini-thesaurus, consisting of a hierarchically structured list of terms on a specific topic or a particular discipline with links to various search engines and web sites. We are exploring the possibility of combining existing methods of information organization with advanced web technology to create an easy-to-use framework for individual web users.

One underlying assumption of this project is that information storage and retrieval methods used by library professionals over the last century have much to offer in the digital environment. Such methods include classification, controlled vocabulary indexing and searching, and sophisticated retrieval strategies. A synthesis of these traditional methods and current web indexing and searching technology may contribute to efforts to overcome the ìanarchy of the Internetî (Lynch, 1997).

A second assumption is that, because of its distributed nature, digital resources may be organized from the bottom up rather than from the top down, as is the case for the organizing tools of the print environment (Chan, 1994). Instead of beginning with the whole universe of knowledge, one may start with more specialized areas of knowledge, building smaller blocks, which may eventually lead to a larger, more comprehensive structure.

Reworking traditional devices of intellectual organization and combining them with modern technology should provide a means of improving organization and management of web resources. Specifically, the traditional information storage and retrieval methods we wish to

††††[8]In this paper, we use the term "Knowledge Class" to represent the concept and "knowledge class or classes" to represent instances of the proposed device.

ISSN: 2324-9773

Lin, X., & Chan, L. (1997). Knowledge Class' - A dynamic structure for subject access on the web. 8th ASIS SIG/CR Classification Research Workshop, 31-40. doi:10.7152/acro.v8i1.12726

capture are (1) the abilities to organize knowledge in a systematic and logical fashion to facilitate browsing, (2) control of synonyms and homonyms typical of a controlled vocabulary in order to improve precision and recall, and (3) devising search strategies to optimize retrieval results.

Classification has been designed for organizing knowledge in a systematic and logical fashion, because knowledge is gained through the association of similar concepts. Information viewed through an organized structure may be easier to perceive and to comprehend. Classification also provides a way to navigate through a large amount of information. Increasingly, popular search engines are employing classification or hierarchical structures to organize web resources.

In the web environment, human indexing of all or most documents using controlled vocabularies is not an option. However, we may explore the control of synonyms and homonyms through search strategies, using controlled vocabularies at the point of retrieval rather than storage, an approach similar to post-coordination. The distinction between the preferred terms and non-preferred terms is important in the manual environment. In the electronic environment, it is possible to include synonymous terms in a strategy, thus shifting synonym control from indexing to searching. In a controlled vocabulary, homonym control is usually achieved through the use of qualifiers, which provide a context for the term in question. Typically, the qualifiers consist of broader terms, or terms from higher levels in the hierarchy. This method may also be adapted to web searching, by including contextual terms in the search strategy. We wish to explore if it is feasible to retain some of the advantages of controlled vocabulary indexing through search strategies.

What we seek to do is to design an interface that is a combination of controlled vocabulary and free-text search engines, which would enable retrieval of documents on the web as well as creation and maintenance of dynamic links to remote resources. It should be flexible enough for users to be able to tailor the scope and the depth of terms stored and the links created. The interface should be equally useful to individuals who wish to develop personalized digital libraries and to information professionals who need to design customized digital libraries focusing on specific subject domains for their clients.

Knowledge Class

Knowledge Class contains two basic components: an organizing framework and an interface for access and retrieval.

Organizing Framework

Basically, the organizing framework is a classified mini-thesaurus, consisting of a hierarchically structured collection of terms on specific topics or particular disciplines, such as investment, the solar system, speech disorders, information science, etc. The terms may be those gathered from existing thesauri or they may be natural language terms based on one's own knowledge or garnered from previous searches. The hierarchical structure may be a branch from an existing classification scheme, or may be built from bottom up by categorizing a collection of terms. The

ISSN: 2324-9773

Lin, X., & Chan, L. (1997). Knowledge Class' - A dynamic structure for subject access on the web. 8th ASIS SIG/CR Classification Research Workshop, 31-40. doi:10.7152/acro.v8i1.12726

emphasis of this framework is on the structure of information and the semantic relationships among terms, topics, branches of subject areas, etc.

Interface for Access, Organization, and Retrieval

In order for the framework to be of use in the web environment, a mechanism capable of linking the hierarchical structure and the component terms to web resources is needed. The mechanism serves as an interactive interface between the user and the terms in the organized framework as well as an interface between the user and the web resources. Through this device, the user can initiate searches by selecting pre-stored terms or search strategies or connect to specific sites previously discovered and stored by clicking on special icons. In the authoring mode, the mechanism will also provide functions to let the user add controlled vocabulary or free-text terms to the hierarchical framework, collect and store new links found during the retrieval process, and expand or modify the hierarchical structure to accommodate additional levels and branches.

At this point, a graphical interface in Java/JavaScript has been developed for this mechanism. The interface is designed with the following capabilities:

- Display of terms in a hierarchical structure representing concepts on various levels Expandable and contractible branches
- Storage of search strategies consisting of terms that may be different from the display terms
- Automatic construction of a search strategy including the term in question and broader terms as qualifiers
- Storage of selected static links to remote resources and related sites or pages
- Dynamic links to multiple search engines such as AltaVista, InfoSeek, Yohoo!, and Lycos, etc.
- Referral links among terms within a knowledge class and potentially among knowledge classes to assist cross-referencing.

Figure 1 shows a display of the knowledge class, "Investment." In this display, the user receives four types of information in four frames. The top right frame provides all six branches of this knowledge class. The top left frame shows details of the selected branch, Forms of investment. Triangular icons before the terms indicate if the terms can be expanded or contracted when the icons are clicked. Global icons after the term "Foreign exchange" and the term "Mortgages" indicate there are hard links to resources related to the terms. Clicking on the icon will lead directly to those resources. Clicking on any of the terms will activate a search. The arrow icons after terms "Mutual funds" and "Stocks" indicate there are cross-references to and from these topics within this knowledge class. Clicking on the arrow icons will go directly to the sections that these cross-references refer to. The largest frame is used to display the retrieval result, in this case, through the selected search engine, AltaVista, after the term "Horses" of the sub-branch "Tangibles" is selected. The actual query used for the retrieval is "Horses investment Tangibles," which is automatically constructed based on the hierarchy. If there is a specifically constructed search query stored (transparent to the user) under this term, the search engine would use the pre-stored search query instead of this automatically constructed query. The small

ISSN: 2324-9773

Lin, X., & Chan, L. (1997). Knowledge Class' - A dynamic structure for subject access on the web. 8th ASIS SIG/CR Classification Research Workshop, 31-40. doi:10.7152/acro.v8i1.12726

bottom left frame shows the currently selected search engine and all available search engines for this knowledge class.

The user can interact with this display in several ways. If the user wants to switch to another branch, he or she can simply select that branch in the top right frame, and the selected branch will then be displayed in the top left frame. The user can also follow the cross-reference links to other branches. If he or she would like to see more details of a particular level, he or she can expand a sub-branch or contract other sub-branches when needed. In the retrieval mode, the user can select different search engines (in the bottom left frame) and use the same query to access a variety of resources covered by different search engines. He or she can also select other related terms to explore related information. Finally, the user can simply construct a new query with a combination of terms displayed in the term list and terms discovered in the retrieved results. If some sites are considered particularly interesting, the user can store hard links under the selected term for future use. A "hard link" is indicated by an icon after the term.

Features of Knowledge Class

The following section discusses the specific features we attempt to implement in Knowledge Class.

Knowledge Class provides a dynamic structure for subject access on the web. Knowledge Class brings related information together in a framework in order to facilitate the interaction between the searcher and the information. When users select a term in a knowledge class, they will be able to perceive semantically related terms through the hierarchy. They will have access to other related information through the built-in cross references. Thus, when a user finds a knowledge class of interest, he or she will have immediate access not only to the specific information being sought, but also to other related information through both "soft" links and "hard" links stored in the knowledge class.

Knowledge Class emphasizes intellectual construction of word/term relationships. Because each knowledge class is designed for a specific subject domain and for individual users, terms in a knowledge class can be carefully selected and constructed by information professionals and can be customized by individual users. The information professional provides a knowledge structure that is based on information organizing principles, and the individual user can modify the structure based on his or her own information needs. Soft links connecting terms in a knowledge class to related information through search engines are not as volatile as hard links such as bookmarks or links to clearinghouse pages. This feature ensures that efforts put in by information professionals have longer lasting usefulness on organizing web resources.

Knowledge Class provides a search query for each display term. Currently, the default search strategy for each term consists of the term itself and all its upper hierarchical terms, plus a ìscope term,î similar to a qualifier in a controlled vocabulary, which helps define the scope of all the terms in this knowledge class. In figure 1, the user selects the term "Horses." Because this is a knowledge class on Investment, the user is certainly interested in horses as tangible investments, rather than other aspects of horses. The query, "horses investment tangibles," reflects this interest and provides an appropriate context.

ISSN: 2324-9773

Lin, X., & Chan, L. (1997). Knowledge Class' - A dynamic structure for subject access on the web. 8th ASIS SIG/CR Classification Research Workshop, 31-40. doi:10.7152/acro.v8i1.12726

The search terms may also be completely different from the display terms, thus giving the designer the flexibility to accurately define a search strategy for any display term. Ineffective search terms may be deleted from the strategy; and other broader terms representing the context may be added, thereby ensuring a certain measure of homonym control. By adding synonyms to the search term, a degree of synonym control can also be achieved.

Furthermore, by allowing search terms to be different from display terms, the device is able to accommodate linguistic variations such as different languages or different dialects of the same language. Display terms and search terms may be in two different languages, as demonstrated in Figure 2, which shows a knowledge class in English and Welsh. Either language may be used in the display and the user may select search terms in the other language. When a term is selected, the search results based on either of the two language terms are very similar. In this example, when the display term "football" is selected in either language, search results based on the selected term are relevant from the point of view of Welsh, i.e., soccer, rather than American football (this is because the term "Welsh" is used as a scope term). Incidentally, this also solves some problems of polysemous terms.

Knowledge Class makes web information access transparent to the user. For those who are not familiar with sophisticated retrieval methods or who lack sufficient computer expertise, assistance in using search engines is often needed either in query construction or in search result displays. Such assistance is built-in with the Knowledge Class interface. The user of a knowledge class often does not need to have specific information such as the URL of the chosen search engines or exact understanding of query construction strategies. He or she only needs to click on a term to bring new information to the screen because each term in a knowledge class has either a pre-stored search query or an automatically constructed search query based on the subject hierarchies. Sophisticated users will always have the option to construct their own queries while maintaining the ability to view and use the terms in the hierarchical list.

Knowledge Class provides a relatively stable information space while maintaining dynamic links to adapt to rapid changes in the digital environment. Users who are interested in a certain topic often desire to re-visit previously found sites related to the topic. If they wished to use the common search engines to go back to sites they visited before, they would have to have stored a bookmark or recall the exact URLs or the queries and search engines they used before. Even then, the previous site may have been changed because of the ever-changing nature of web information resources. If they use a knowledge class to re-visit the sites they will only need to interact with a familiar structure to keep in contact with the changing environment. As we continue to improve the interface, users will also be able to customize the familiar space to reflect their information needs. They will be able to store new links associated with existing terms, add new terms to the knowledge class by drag-and-drop, and expand the hierarchy to accommodate additional levels. As their interest and their understanding of the topic grows, they will find that they can also make the knowledge class "grow."

Summary

ISSN: 2324-9773

This research project attempts to devise a new method of knowledge organization in the distributed environment. It investigates how traditional information organizing concepts, methods, and tools may be applied to the digital environment.

Knowledge Class is designed to be both an information organizing device and an information access tool. Individually, each knowledge class is an organized entity dynamically linked to web information resources. Collectively, instances of knowledge classes can be assembled through classification hierarchies to form a conceptual infrastructure for organizing and managing web resources. The concept of Knowledge Class is built on traditional devices of knowledge organization and access, specifically classification principles and controlled vocabulary retrieval.

Principles and practice of online searching, representing the cumulative wisdom and experiences of information professionals, are adopted and modified for use in the web environment. Finally, the interface is built with recent web technology and based on human computer interaction principles. It is hoped that this research project will have both theoretical and practical values.

A demo of Knowledge class can be seen at http://lislin.gws.uky.edu/kc/index.html. Several examples of knowledge classes are available for viewing. These examples are based on a working prototype developed in Netscape's javaScript. A Java version of Knowledge Class is currently under development and will soon be put online. Currently, we are planning to conduct user studies on how subject specialists use and interact with knowledge classes in their subject domains. We hope that these "real users" will provide valuable feedback about the concept of Knowledge Class we are developing.

Acknowledgement

We would like to thank students in the class "Content analysis of web services" that we taught in spring, 1997 at the School of Library and Information Science, University of Kentucky. They have contributed to the development and testing of Knowledge Class. We would also like to thank anonymous reviewers for their detail and excellent comments. This research was supported by a research committee grant from University of Kentucky.

References

Chan, Lois Mai. (1994). Cataloging and classification: An introduction. 2nd ed. New York : McGraw-Hill.

Koster, Martijn (1995). Robots in the web: threat or treat? ConneXions, 9(4), April 1995. [web page] http://info.webcrawler.com/mak/projects/robots/threat-or-treat.html

Lynch, Clifford (1997). Searching the Internet. Scientific American. 276(3), 52-56, March 1997.

Lin, X., & Chan, L. (1997). Knowledge Class' - A dynamic structure for subject access on the web. 8th ASIS SIG/CR Classification Research Workshop, 31-40. doi:10.7152/acro.v8i1.12726

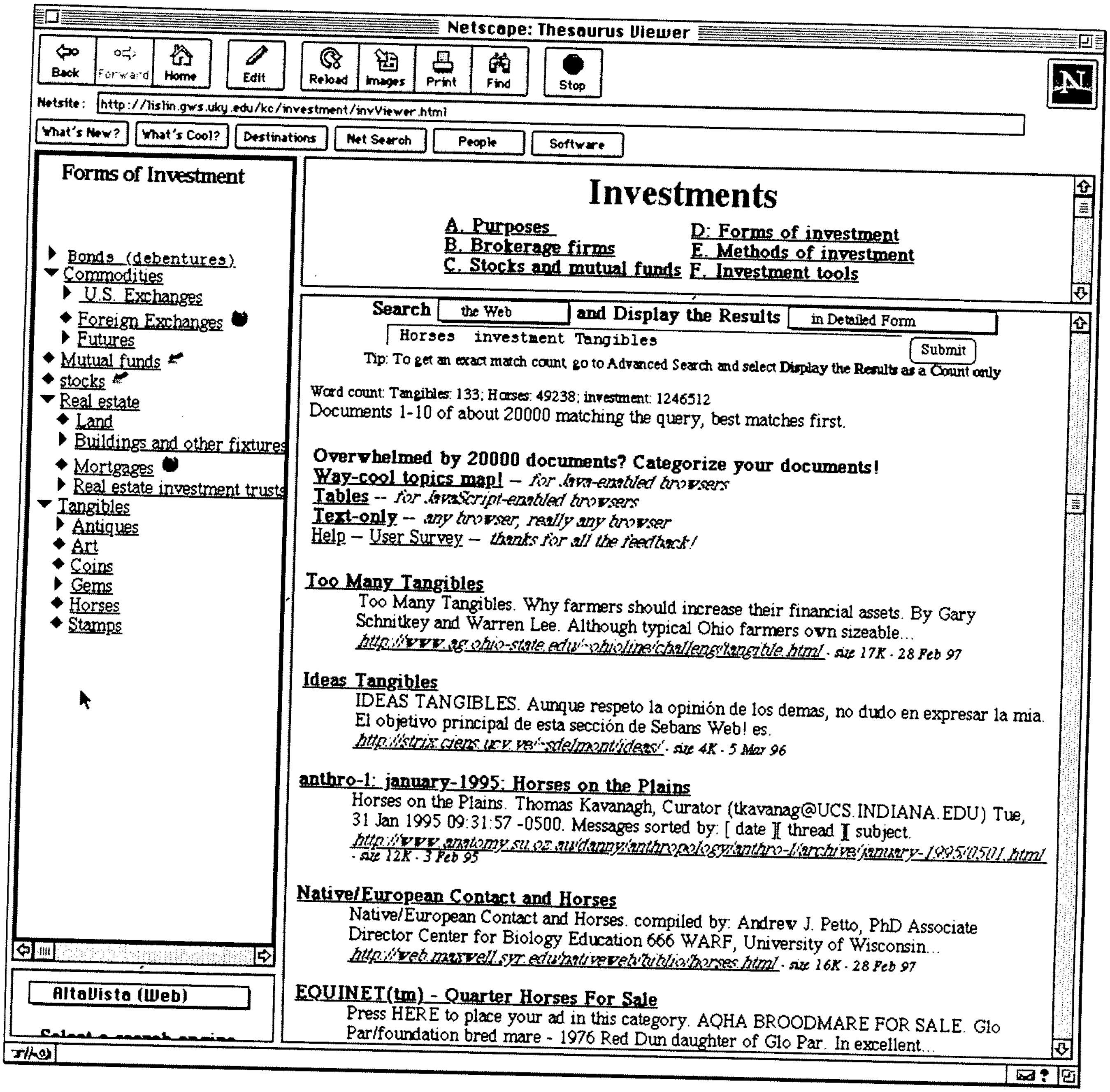
McKiernan, G. (1997). Beyond Bookmarks: A Review of Frameworks, Features, and Functionalities of Schemes for Organizing the Web. Internet Reference Services Quarterly 2(1/2) 1997.

Notess, Greg R. (1995). Comparing Commercial WWW Browsers. Online, 19(3), 43-49. May-June 1995.

Weibel, Stuart. (1995). Metadata: The Foundations of Resource Description. D-lib Magazine. [web page] http://www.dlib.org/dlib/July95/07weibel.html.

ISSN: 2324-9773

Lin, X., & Chan, L. (1997). Knowledge Class' - A dynamic structure for subject access on the web. 8th ASIS SIG/CR Classification Research Workshop, 31-40. doi:10.7152/acro.v8i1.12726



Netsite: http://lislin.gws.uky.edu/kc/investment/invViewer.html

Forms of Investment

▶ Bonds (debentures)
▼ Commodities
  ▶ U.S. Exchanges
  ◆ Foreign Exchanges ●
  ▶ Futures
◆ Mutual funds
◆ stocks
▼ Real estate
  ◆ Land
  ▶ Buildings and other fixtures
  ◆ Mortgages ●
  ▶ Real estate investment trusts
▼ Tangibles
  ▶ Antiques
  ◆ Art
  ◆ Coins
  ▶ Gems
  ◆ Horses
  ◆ Stamps

AltaVista (Web)

Select a search engine

Investments

A. Purposes
B. Brokerage firms
C. Stocks and mutual funds
D: Forms of investment
E. Methods of investment
F. Investment tools

Search the Web **and Display the Results** in Detailed Form

Horses investment Tangibles

Tip: To get an exact match count, go to Advanced Search and select **Display the Results as a Count only**

Word count: Tangibles: 133; Horses: 49238; investment: 1246512
Documents 1-10 of about 20000 matching the query, best matches first.

Overwhelmed by 20000 documents? Categorize your documents!
Way-cool topics map! -- *for Java-enabled browsers*
Tables -- *for JavaScript-enabled browsers*
Text-only -- *any browser, really any browser*
Help -- User Survey -- *thanks for all the feedback!*

Too Many Tangibles
Too Many Tangibles. Why farmers should increase their financial assets. By Gary Schnitkey and Warren Lee. Although typical Ohio farmers own sizeable...
http://www.ag.ohio-state.edu/~ohioline/challeng/tangible.html - size 17K - 28 Feb 97

Ideas Tangibles
IDEAS TANGIBLES. Aunque respeto la opinión de los demas, no dudo en expresar la mia. El objetivo principal de esta sección de Sebans Web! es.
http://strix.ciens.ucv.ve/~sdelmont/ideas/ - size 4K - 5 Mar 96

anthro-l: january-1995: Horses on the Plains
Horses on the Plains. Thomas Kavanagh, Curator (tkavanag@UCS.INDIANA.EDU) Tue, 31 Jan 1995 09:31:57 -0500. Messages sorted by: [date][thread][subject.
http://www.anatomy.su.oz.au/danny/anthropology/anthro-l/archive/january-1995/0501.html - size 12K - 3 Feb 95

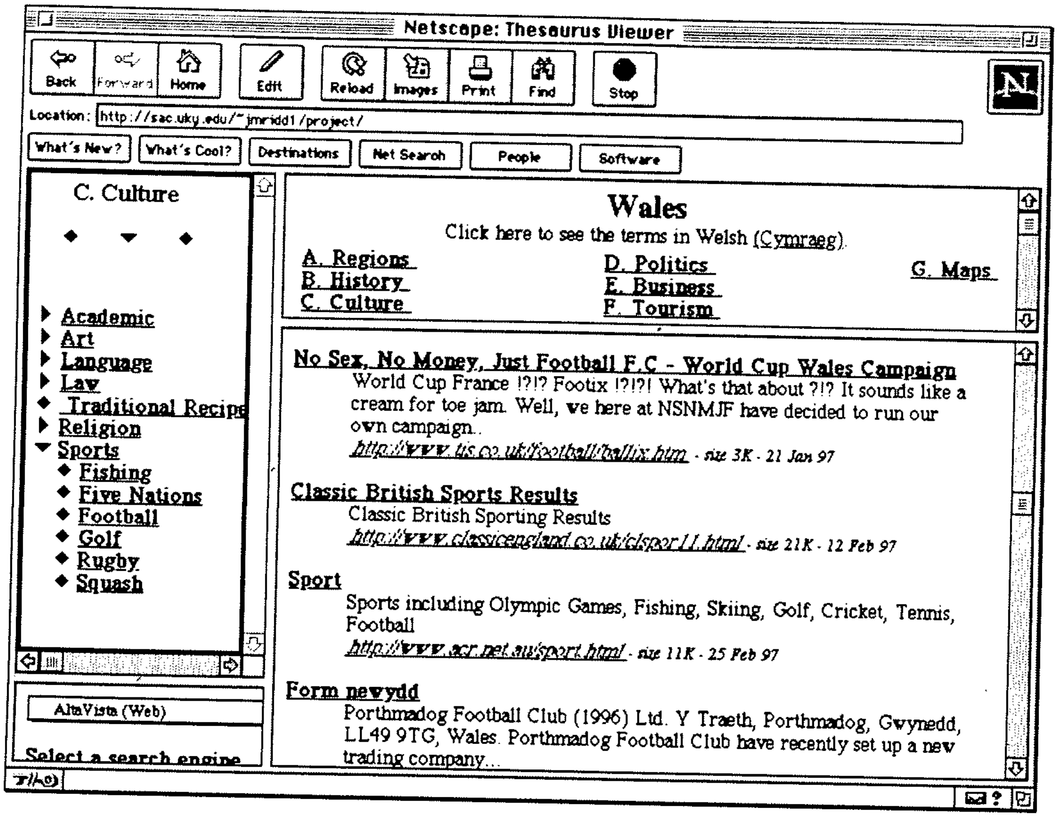
Native/European Contact and Horses
Native/European Contact and Horses. compiled by: Andrew J. Petto, PhD Associate Director Center for Biology Education 666 WARF, University of Wisconsin...
http://web.maxwell.syr.edu/nativeweb/biblio/horses.html - size 16K - 28 Feb 97

EQUINET(tm) - Quarter Horses For Sale
Press HERE to place your ad in this category. AQHA BROODMARE FOR SALE. Glo Par/foundation bred mare - 1976 Red Dun daughter of Glo Par. In excellent...

Fig.1 A screen dump of Knowledge Class Viewer

ISSN: 2324-9773

Lin, X., & Chan, L. (1997). Knowledge Class' - A dynamic structure for subject access on the web. 8th ASIS SIG/CR Classification Research Workshop, 31-40. doi:10.7152/acro.v8i1.12726



Location: http://sac.uky.edu/~jmriddl/project/

C. Culture

◆ ▼ ◆

- ▶ Academic
- ▶ Art
- ▶ Language
- ▶ Law
- ◆ Traditional Recipe
- ▶ Religion
- ▼ Sports
  - ◆ Fishing
  - ◆ Five Nations
  - ◆ Football
  - ◆ Golf
  - ◆ Rugby
  - ◆ Squash

Wales

Click here to see the terms in Welsh (Cymraeg).

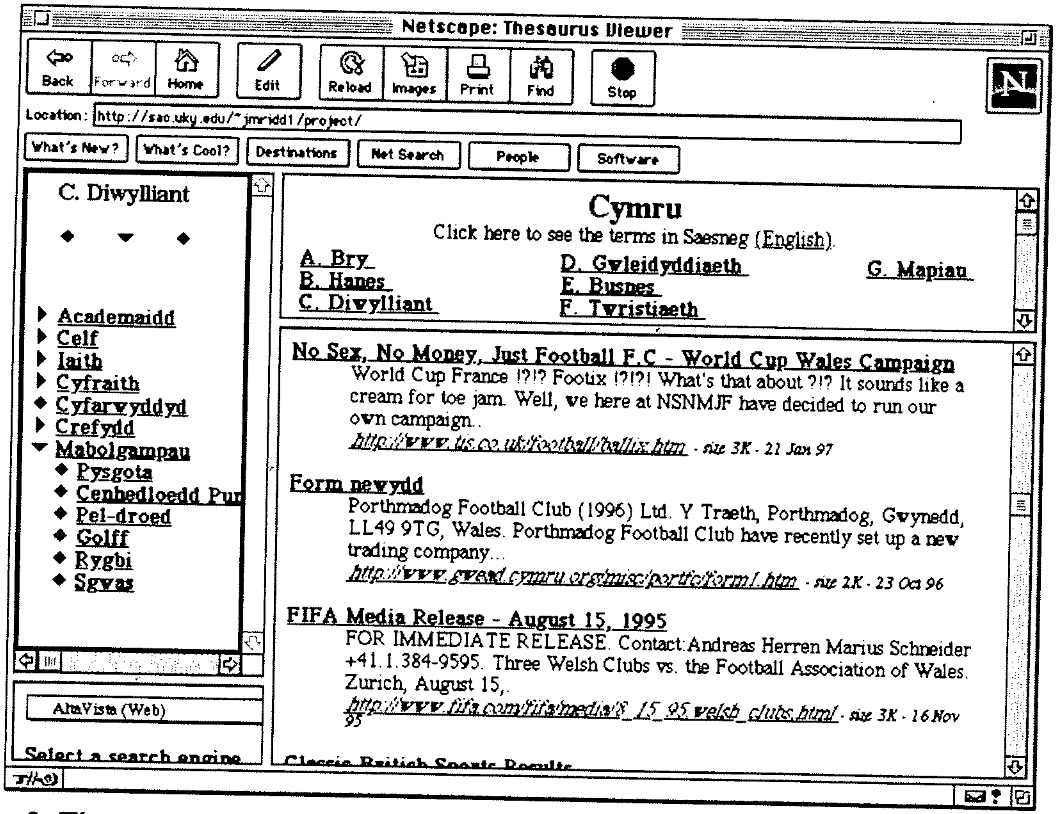
A. Regions	D. Politics	G. Maps
B. History	E. Business	
C. Culture	F. Tourism	

No Sex, No Money, Just Football F.C - World Cup Wales Campaign
World Cup France !?!? Footix !?!?! What's that about ?!? It sounds like a cream for toe jam. Well, we here at NSNMJF have decided to run our own campaign..
http://www.tis.co.uk/football/ballix.htm - size 3K - 21 Jan 97

Classic British Sports Results
Classic British Sporting Results
http://www.classicengland.co.uk/clspor11.html - size 21K - 12 Feb 97

Sport
Sports including Olympic Games, Fishing, Skiing, Golf, Cricket, Tennis, Football
http://www.acr.net.au/sport.html - size 11K - 25 Feb 97

Form newydd
Porthmadog Football Club (1996) Ltd. Y Traeth, Porthmadog, Gwynedd, LL49 9TG, Wales. Porthmadog Football Club have recently set up a new trading company...



Location: http://sac.uky.edu/~jmriddl/project/

C. Diwylliant

◆ ▼ ◆

- ▶ Academaidd
- ▶ Celf
- ▶ Iaith
- ▶ Cyfraith
- ◆ Cyfarwyddyd
- ▶ Crefydd
- ▼ Mabolgampau
  - ◆ Pysgota
  - ◆ Cenhedloedd Pum
  - ◆ Pel-droed
  - ◆ Golff
  - ◆ Rygbi
  - ◆ Sgwas

Cymru

Click here to see the terms in Saesneg (English).

A. Bry	D. Gwleidyddiaeth	G. Mapiau
B. Hanes	E. Busnes	
C. Diwylliant	F. Twristiaeth	

No Sex, No Money, Just Football F.C - World Cup Wales Campaign
World Cup France !?!? Footix !?!?! What's that about ?!? It sounds like a cream for toe jam. Well, we here at NSNMJF have decided to run our own campaign..
http://www.tis.co.uk/football/ballix.htm - size 3K - 21 Jan 97

Form newydd
Porthmadog Football Club (1996) Ltd. Y Traeth, Porthmadog, Gwynedd, LL49 9TG, Wales. Porthmadog Football Club have recently set up a new trading company...
http://www.gwead.cymru.org/misc/portfc/form1.htm - size 2K - 23 Oct 96

FIFA Media Release - August 15, 1995
FOR IMMEDIATE RELEASE. Contact:Andreas Herren Marius Schneider +41.1.384-9595. Three Welsh Clubs vs. the Football Association of Wales. Zurich, August 15,.
http://www.fifa.com/fifa/media/8_15_95welsh_clubs.html - size 3K - 16 Nov 95

Classic British Sports Results

Fig.2 The same knowledge class in two languages, English and Welsh. When term "football"is selected, the search results are similar.

ISSN: 2324-9773