A Study on Implementing Topic Maps-based Web search services in junior high school library of WuGu

Wei-Fu Hou
Head of MIS Office
WUGU junior high school
Taipei Taiwan

Web services have become well known in the school library, yet adoption of them with in the school setting still little have been down, especial in junior high school library. In this paper, we develop a Grade 1-9 Curriculum Web resources search services by using Web services, Grade 1-9 Curriculum, Topic Maps, XTM and relational database management system (RDBMS). The Grade 1-9 Curriculum was transferred into a Topic Maps to representation domain-specific knowledge model and to connect to related resources. This domain-specific knowledge of a topic is exploited by the Web search services to suggest appropriate search queries and retrieval adequate information from the internet. Finally, these functions are integrated into the WUGU junior high school library system and provide an efficient way for users to access heterogeneous information.

Introduction

In the decade past, the client-server network architectures have let many enterprises take a great leap in benefit. Today, Web services have come to maturity with two well-established technologies: service-oriented architectures (SOA) and proven Internet protocol and standard (HTTP, POST, GET, REST, etc.). As Gartner's Dean Lombardo (2001) says, "By 2005, the aggressive use of web services will drive a 30% increase in the efficiency of IT development projects." At first glance, Web services are simple and easily reused, so long as the conformant with its standards and description. Yet adoptions of Web services technologies with in the junior high school library still little have been down (Breeding, 2006). One of the reasons is what and how Web Services can be implementing in library context.

What Web Services can be implementing in library context？NISO Web Services and Practices Working Group (2006) has outlined five categories as followings :(1) Discovery Services ;(2) Locate Services; (3) Requesting Services ;(4) Delivery Services ;(5) Common Services. When considering the full range of these Web services-related activities in junior high school library, it becomes clear that dealing with Information retrieval (IR) functions is a critical issue in connection with Web services but how junior high school library move from isolated digital collections to interoperable digital collections is also a major challenge.

This paper endeavors to present an information retrieval functions for Grade 1-9 Curriculum(Department of Elementary Education, 2006) Web resources, which was adopted as Web services. A key point behind Web services is that these frequently used functions can be implemented once and offered to Integrated Library System. Moreover, through the

semantic tagging provided by local repository of ontological data, this system led users can feel easily retrieval precise relevant information across networks and between organizations, and allow dispersed digital collections interoperable.

Literature review

The field of Information retrieval research has made more progress for years. However, the increasing volume of heterogeneous collection of information resources with minimal selection, organization, and retrieval standards available on the web makes IR a tedious and difficult task. Therefore, the Semantic Web is addressed this challenge by trying to make the Web data machine-understandable, and by automatically extracting the useful knowledge hidden in these data from World Wide Web Consortium.

In this section, we reviewed related works about information retrieval and the format about semantic information, which describing resources and establish relationships among them.

First, the field of information retrieval research is commonly distinguished: text retrieval, data retrieval and knowledge retrieval (Huang, 2005). But comparing with the three types of information retrieval, knowledge retrieval has own features. The most important difference between knowledge information and other Web information is that knowledge information has rigid semantic structure. How to improve quality and precision of search by semantic analysis is key factor in designing search engine.

Many studies in IR consider keyword as important factor. For example, Chen (2002) introduces a Hyperlink and Keyword Frequency Analysis (HKFA) search engine in his IR research. However, the HKFA search engine only considers algorithms based on Hyperlink and Keyword frequencies as web page level and dose not check semantic of search objects. Lee (2005) and Hu (2004) also use the Google API based on keyword to search English and Chinese resources on the WWW. But the most problem about keyword based search is that user may provide inappropriate keywords and bring about related and unrelated query results. In order avoid the problem, another approach is proposed by Lu (2005). Lu designs an agent that exploits users’ search histories, Web page’s theme, and links not keyword for finding related Web pages to help English learning. Unlike common search engines, the system can help user find Web resources easily without ever having to provide keyword. Clearly, above these studies do not change essence of traditional search engines, which were link-based or content-based at all.

Taxonomy can improve IR efficiently was proposed by Lin (2002). The author proposed a digital library that is designed based on the schema of Grade 1-9 Curriculum can help us get useful resources and his study has indicated that the taxonomy is beneficial in improving the searching effectiveness for students. However taxonomy only provide a vocabulary of terms and simple relationships between these terms and the relationships they express are not as rich as the ones provided by RDF of Topic Maps (TM) and consequently by ontologies. Huang developed an inference-based intelligent search engine based on semantic web technology and users' background knowledge to assist literatures retrieval. And the result has revealed ontology based search engine can efficiently improve searching performance than traditional keywords search. In order promote the efficiency of IR, his framework of digital library ontology also added the description of user's characteristic, including job, interest, major and location. While these users' background knowledge was

inadequate in junior high school context, we suppose not the job, major and location but interest and grade will influence the user's behaviour of Web search. However, the framework of digital library ontology was built in laboratory which only contains 95 copies of PDF format document. Compare with the WWW, the content match mechanism and subject taxonomy by manual are another challenge.

At the same time, Chiu (2004) also proposed a semantic web technology based search engine, which was a RDF-based mechanism integrated with word sense disambiguation technique to semantically index and retrieve Web page on the WWW. The study has three features than others. First, it introduces WordNet, which was a lexical database for computational linguistics and natural language processing to enhance semantic discrimination. Second, It not only use relation database to store retrieved Web page based on RDF but also provide search related data management mechanism such as keyword, Synset, Address and REM. However, the author does not evaluate the performance of the RDF-based Search Engine system for his research by any methods yet such as recall and precision. Table 1 shows the two RDF-based search engine researches in summary.

Table 1. RDF-based IR researches

	Fu-Ming Huang (2005)	Mao-Shuen Chiu (2004)
Ontology technology	RDF, OWL	RDF
Relation Database	No	Yes, MySQL
User profile(knowledge)	job, interest, major and location	User request (keyword)
Evaluate	Recall, Precision	No
Search target	Digital Library(95 copies of PDF format)	WWW
Features	1.Description Logic System 2.extract metadata from documents	Semantic filter
Metadata	Predefined.	Not defined

Similar to semantic web technology, Topic Maps (Biezunski, 2001; Pepper, 2002) is another metadata representation format. Although Topic Maps does not appear on the Semantic Web stack (Berners-Lee, 1998) but Topic Maps would be at the same level as RDF in syntaxes, data models, and constraints (Garshol, n.d.). On the other hand, both standards aim to represent connections between information objects and can encode metadata, among other things. In figure 1, the left side row is Topic Maps constructs and the right side row is RDF property.

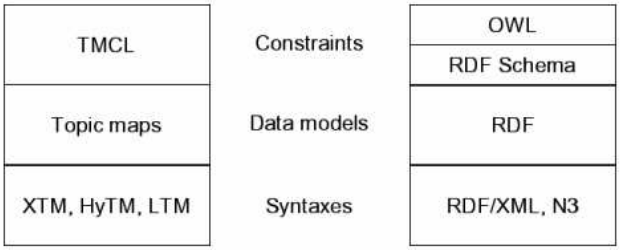

Figure 1. Mapping RDF property to Topic Maps construct.

Besides RDF based IR researches, Chou (2004) combined Topic Maps with data-mining technologies to get more useful and relevant information on the internet and provide a more semantic structure result for users. However, this study does not much mention about IR but proposed a method of producing domain-specific repository. Similar research was done by Hsu (2004). Hsu adopts the Topic Maps technologies in design event-based on-line repository. Meanwhile, in his study, a browsing and searching capability of web-based user interface is available for users. Table 2. Show the two Topic Maps-based IR researches in summary.

Table 2. Topic Maps-based IR researches

	Chung-Chiang Chou (2004)	Teng-chieh Hsu (2004)
Ontology technology	Topic Maps	Topic Maps
Relation Database	Yes, MySQL	Yes, MySQL
Evaluate	Yes, quantity	Yes, questionnaires
Search target	WWW	WWW
Features	Follow XTM specific	Visual Navigation
Limited		Not follow XTM specific

In summary, these IR related studies have indicated that many techniques and operations such as taxonomy, ontology, TM, RDF, OWL and XTM can provide efficiently, useful and practical search service. However, the field of IR researches about domain-appropriate knowledge or representation scheme for children and youngsters are still few in Taiwan.

Besides, in study of elementary schoolteachers' searching behaviour on Learning Fuelling Station has revealed that schoolteachers use the Internet prior to design their lesson plans (Chen, 2005). On the other hand, the need of searching the Internet resources is important part of school learning activities. In Taiwan, Many resources about Grade 1-9 Curriculum on the WWW are still infoglut.

Although RDF is adopted by many researchers, yet Garshol (2004) have demonstrated taxonomies, thesauri, ontology or faceted metadata can be easily be integrated into Topic Maps. Meanwhile, Topic Maps is able to manage related links in dynamic manner and could be used to disambiguate distinct concepts with the same term (homonyms) based on their context in the Topic Maps. According to our study purpose and TM's advantage, this paper would exploit the Topic Maps related concept and technology for our study tools.

Research purpose and method

Research purpose

In this paper, we exploit Yahoo Web Search API (Yahoo, 2006), Grade 1-9 Curriculum, Topic Maps, XTM (Garshol & Moore, 2006) and the RDBMS (Graauw, 2003) perspective to facilitate the information retrieval for Grade 1-9 Curriculum domain resources. Moreover, we demonstrate how these functions can be offered as Web services.

Research method

Grade 1-9 Curriculum Ontology

We use Top Maps to represent ontology. The Topic Maps-based Grade 1-9 Curriculum ontology enables us to agree upon the meaning of terms used in domain-specific knowledge. To structure the field we use the Grade 1-9 Curriculum taxonomy and have been indexing resources in the concepts of this ontology. Figure 2 is Grade 1-9 Curriculum of WUGU junior high school.

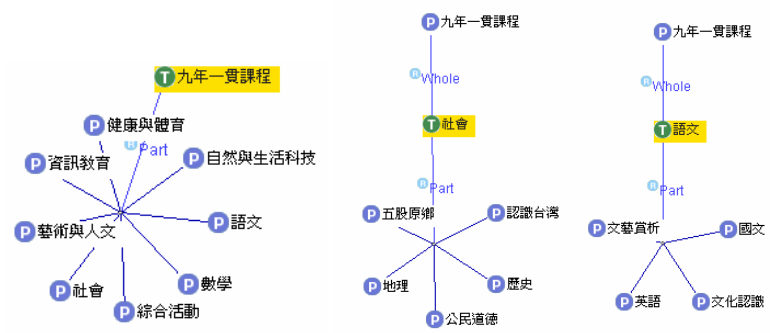

Figure 2. Grade 1-9 Curriculum ontology of WUGU junior high school

The Grade 1-9 Curriculum ontology as it was presented depicts the contents and characteristics of information and the construction of domain knowledge, and further integrates and classifies them. Table 3 shows the steps of construct Grade 1-9 Curriculum ontology by Topic Maps (Lee, 2002).

Table 3. The steps of construct Grade 1-9 Curriculum Ontology

Step	Topic Maps
1	Identify the Learn Areas of Grade 1-9 Curriculum and scope of Topic Maps
2	Acquisition of Grade 1-9 Curriculum's Learn Areas knowledge
3	Define Grade 1-9 Curriculum's Topic、Association、Occurrence
4	Merge Map
5	Encoding by XTM
6	Test、Evaluation

Grade 1-9 Curriculum Repository

Grade 1-9 Curriculum repository is specific using Grade 1-9 Curriculum ontology. Topic, Occurrence and Association are the fundamental components of the Grade 1-9 Curriculum repository. These components constitute metadata that describe the contents of the resources. In other words, Grade 1-9 Curriculum repository contains domain expert-specified description of information resources and serves as metadata for these resources. To allow Grade 1-9 Curriculum repository can be managed, the Grade 1-9 Curriculum repository is store in a RDBMS (see Figure 3). Besides, we could also publish Grade 1-9 Curriculum repository for other school library on their Websites as a web services.

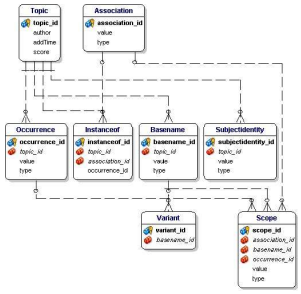

Figure 3. The Grade 1-9 Curriculum repository schema.

The initial step for the Grade 1-9 Curriculum repository is that how to determined the Grade 1-9 Curriculum topics, association and occurrence. This may involve techniques like machine learning, Web mining and IR. In order to extract metadata from the actual Web resources, we combine Yahoo Web search API with the keyword or phrase which provide by domain experts as a tool. Then, a Web-robot traverse the Web creates, and stores the metadata entities in a RDBMS. Once Grade 1-9 Curriculum repository is done, not only covers a distributed set of domain-specific information over the Web but also provides useful, efficiently and practical Web search services.

User Profiles

As mentioned before, user profiles can enhance the efficiently query results of IR. Thus, User profiles are also created and stored in a RDBMS. User profiles are composed of user preferences and user knowledge. User preferences allow each student to specify his/her preferences about personal information such as age, grade, interest, favourite, and so on. User knowledge maintains knowledge of users on grade 1-9 curriculum in terms of detail levels as well a navigational history information for the users. Once a user clicks some links which return by user request, the URL of the information that the user visits, the first and last visit dates, the keyword, and the visit frequency for the information are directly written to the user profiles table.

Web Search Service Architectures

We employ Grade 1-9 Curriculum ontology in Topic Maps model and regard Topic Maps as the reference structure. The system (see Figure 4.) mainly deployed on .NET Framework, and more specifically, by ASP.NET, SQL 2005 and Windows 2003 server.

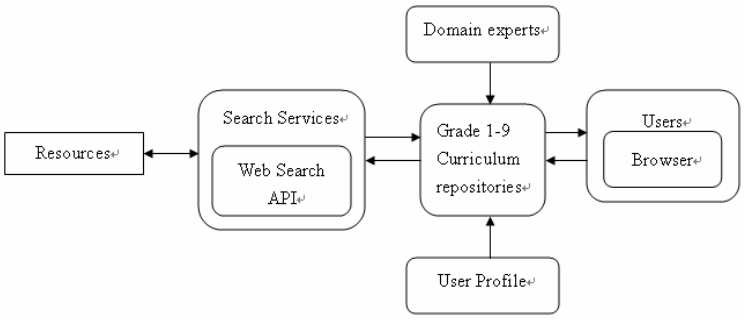

Figure 4. Topic Maps-based Web Search Services for Grade 1-9 Curriculum

Evaluate

We conducted a standard information retrieval task, running queries against engines, pooling the results for each query, and using grade 7 students(N=30) to complete them. To compare with our search engine, a general search engine (Yahoo) was selected. Table 4 lists the search engines included in our study.

Table 4. The search engines included in the study.

Engine	URL	Notes
Yahoo	http://tw.yahoo.com/	General knowledge search
Topic Maps-based Grade 1-9 Curriculum Web search services	http://localhost/lib2/person/MetaSearch_Yahoo.aspx	Grade 1-9 Curriculum domain knowledge search

Considering search engines performance and students' knowledge, we measured 2 queries, relating to Grade 1-9 Curriculum domain topics. Finally, we use descriptive statistics to describe and summarize the results.

Results

The result of this paper is divided into two parts. In the first part, we explain the Web services that provided by the system. In the second part, we discuss the result of domain-specific search tasks.

Web services

In our system, we developed and deployed four Web service methods. Each of these services exposes a number of methods which can be invoked either by a SOAP client or by simple HTTP page requests. These Web services are described below.

Get topics by name

The function which gets topics by name retrieves the requested topic and the service returns an XML document which contains each of the topics that have at least one name that matches the name parameter (see Figure 5). The method has following parameters: (1) Domain: the knowledge domain for Grade 1-9 Curriculum. (2) Name: the name for the requested topic.

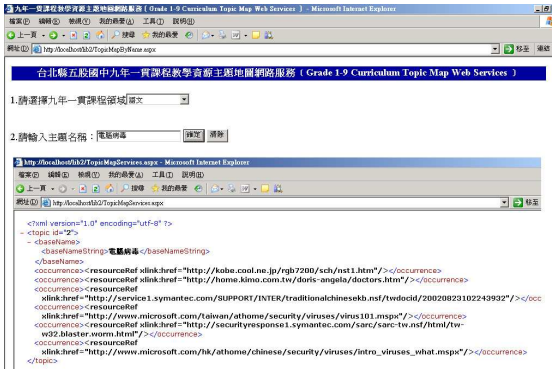

Figure 5. Get topics by name result

Get topics by type

The function which gets topics by type function retrieves the requested topic and the service returns an XML document which contains all topics of a specified type (see Figure 6). The method has following parameters: (1) Domain: the knowledge domain for Grade 1-9 Curriculum. (2) Type: the type for the requested topic.

Figure 6. Get topics by type

Get Topic Types by Topic Maps

The function which gets topic types by the name of Topic Maps retrieves the requested topic types and the service returns an XML document which contains a collection of topic types in the named Topic Maps (see Figure 7). The method has following parameters: (1) Domain: the knowledge domain for Grade 1-9 Curriculum. (2) Name: the name for the requested Topic Maps.

Figure 7. Get topic types by Topic Maps result

Topic Maps-based Grade 1-9 Curriculum Web search services

The function which gets information resources by the name of topic retrieves the requested Grade 1-9 Curriculum domain resources and the service returns collection resources of a specific domain (see Figure 8). The method has following parameters: (1) Domain: the knowledge domain for Grade 1-9 Curriculum. (2) Name: the name for the requested topic.

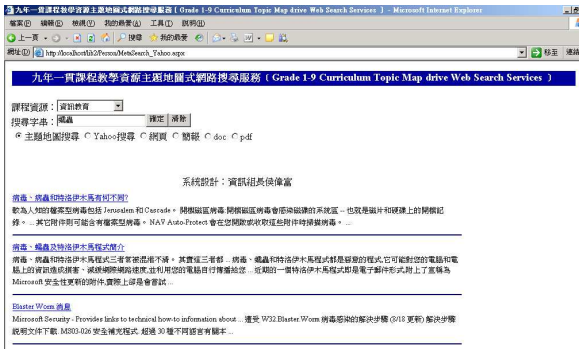

Figure 8. Topic Maps-based Grade 1-9 Curriculum Web search services results

Grade 1-9 Curriculum domain-specific search results

Table 5 shows the comparison of correct answer between Yahoo search and Topic Maps-based Grade 1-9 Curriculum Web search services.

Table5. The comparison of search engines search results (N＝30)

	Query 1	Query 2
Yahoo search	34.62%	16.67%
Topic Maps-based Grade 1-9 Curriculum Web search services	65.38%	83.33%

The result shows the Topic Maps-based Grade 1-9 Curriculum Web search services is better than general search engine. The meaning of this evaluation show that ontology based search is still more effective than traditional keywords search.

Conclusion and suggestions

Conclusion

This paper has attempted to demo practical implementation issues that arise while defining the Web Services strategy for junior high school library. It also demonstrated that search services can be successful access Grade 1-9 Curriculum resources by Topic Maps. To sum up the research results, the conclusions are described as follows:

1. Topic Maps can represent the relationships that cannot be held in domain knowledge schema. By reasoning the association described in the topics, the system can finds out the additional implicit meaning of the user query. As such, it can discover more precise and necessary information regarding the user query.

2. In search tasks, we found that it still have many potential factors influence ushers' searching behaviours, such as motive and knowledge. So the preceding analyses of the factors that would influence searching are indispensable.

3. The prototype system implemented in this paper indeed enhances the interoperability between the heterogeneous information sources. Through the function of

Web services, users or junior high school libraries can get the resources and integrated information from different heterogeneous information source.

Suggestions

The prototype system for Grade 1-9 Curriculum Web search services is trying to prove the Topic Maps-based method we provide and get a good result. Due to time and resource restrictions, this study still has a number of areas to improve in the future. For example, the automatic or semi-automatic create topics and merge maps method may need further study. Also evaluate and measure the Web search services more precise can improve the efficiency and effectiveness of the functions.

Acknowledgements

I would like to thank Dr. Chao-Chen Chen, the three anonymous reviewers and my friends for their valuable suggestions and comments. These have helped me improve the contents of the paper.

References

Berners-Lee, Tim. (1998). Semantic Web Road map. Retrieved November 3, 2006, from the World Wide Web:http://www.w3.org/DesignIssues/Semantic.html.

Biezunski, M. (2001). Topic maps at a glance. Retrieved November 5, 2006, from the World Wide Web: http://www.infoloom.com/tmsample/bie0.htm.

Breeding, M. (2006). Web Services in the Library Environment. Library Technology Reports, 42(3), p36-46. Retrieved November 5, 2006 from the World Wide Web: http://web.ebscohost.com/.

Chen, Chih-Ho.（2002）.A Study on Implementation for Search Engine Based on Hyperlink and Keyword Frequency Analysis. (Master thesis, National Pingtung University of Science and Technology, 2002). Electronic Theses and Dissertations System, 090NPUST396014.

Chen, Shih-Chuan. (2005). A Study of Elementary Schoolteachers' Searching Behaviour on Learning Fuelling Station. (Master thesis, Nation Taiwan University, 2005). Electronic Theses and Dissertations System, 094NTU05447011.

Chiu, Mao-Shuen. (2004).The Research and Implementation of Semantic Based RDF Tagging and Webpage Searching Web Service. (Master thesis, Tamkang University, 2004). Electronic Theses and Dissertations System, 093TKU05392007.

Chou, Chung-Chiang. (2004). Study on the Application of Topic Maps in Knowledge Mining. (Master thesis, Nation Taiwan University, 2004). Electronic Theses and Dissertations System, 093NTU05345004.

Department of Elementary Education. (2006). Grade 1-9 Curriculum. Retrieved November 4, 2006, from the World Wide Web: http://www.edu.tw/EDU_WEB/EDU_MGT/EJE/EDU5147002/9CC/9CC.html.

Garshol, L. M. (n.d.). Living with topic maps and RDF. Retrieved November 10, 2006, from the World WideWeb: http;//www.ontopia.net/topicmaps/materials/tmrdf.html.

Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic Maps! Retrieved November 10, 2006 from the World Wide Web: http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html.

Garshol, L. M., Moore, G. (2006). Topic Maps-XML Syntax. Retrieved November 3, 2006, from the World Wide Web: http://www.isotopicmaps.org/sam/sam-xtm/.

Graauw, M. D. (2003). Using Topic Maps to Extend Relational Databases. Retrieved November 9, 2006, from the World Wide Web: http://www.infoloom.com/tmsample/bie0.htm.

Hu, Chich-Hsiang. (2004). Using Meta-Search for Searching Chinese Sentences. (Master thesis, Yuan Ze University, 2004). Electronic Theses and Dissertations System, 093YZU00396048.

Hsu, Teng-chieh. (2004). Applying Topic Maps theory to construct knowledge indexes on event-based on-line news retrieval research. (Master thesis, National Yunlin University of Science and Technology, 2004). Electronic Theses and Dissertations System, 093YUNT5396035.

Huang, Fu-Ming. (2005). Intelligent Search Engine with Semantic Web Technology. (Master thesis, National Central University, 2005). Electronic Theses and Dissertations System, 093NCU05726003.

Lee, Chia-Hao. (2002). Design and Implementation of A Bio-Ontology Topic Maps. (Master thesis, National Pingtung University of Science and Technology, 2002). Electronic Theses and Dissertations System, 090NPUST396003.

Lee, Kai-Te. (2005). Using Google API for Searching English Sentences. (Master thesis, Yuan Ze University, 2005). Electronic Theses and Dissertations System, 093YZU00396045.

Lin, Chih-Neng. (2002). The Study on Retrieval of Networked Learning Materials. (Master thesis, National Taiwan Provincial Tainan Teachers College, 2002). Electronic Theses and Dissertations System, 090NTNTC395002.

Lombardo, D., gartner.com (2001). The Hype Is Right: Web Services Will Deliver Immediate Benefits. Retrieved November 3, 2006, from the World Wide Web: http://www.gartner.com/resources/101500/101571/101571.pdf.

Lu, Ming-Feng. (2005). The Design of an Agent for Finding Related Web Pages and its Application to English Learning. (Master thesis, Tamkang University, 2005). Electronic Theses and Dissertations System, 094TKU05392035.

NISO Web Services and Practices Working Group. (2006). Best Practices for Designing Web Services in the Library Context. Retrieved November 3, 2006, from the World Wide Web:http://www.niso.org/standards/resources/rp-2006-01.pdf.

Pepper, S. (2002). The TAO of Topic Maps. Retrieved November 3, 2006, from the World Wide Web: http://www.ontopia.net/topicmaps/materials/tao.html.

Yahoo. (2006). Web Search APIs from Yahoo! Search Web Services. Retrieved December 10, 2006, from the World Wide Web: http://developer.yahoo.com/search/index.html.