December 2002

DESIGN SPACE OF PERSONALIZED INDEXING: ENHANCING SUCCESSIVE WEB SEARCHES FOR TRANSMUTING INFORMATION PROBLEMS

Shin-Jeng Lin
Le Moyne College

Design Space of Personalized Indexing: Enhancing Successive Web Searches for Transmuting Information Problems

Shin-jeng Lin
Le Moyne College
linsj@mail.lemoyne.edu

Abstract

Successive searches are very common search behavior in online shopping, Intranet information retrieval and other ordinary information seeking contexts. This study unveils the design space of Personalized Indexing (PI), a tool that supports successive searches for information problems that evolve over sessions. An empirical user study on a representative task of transmuting information problems in the Web context was conducted to evaluate PI. The preliminary result of this user study is briefly reported, confirming that PI does support the searchers well in the problematic areas of the information seeking process it intends to address.

Keywords: Personalized indexing, interface design, successive searches, information gathering, information seeking

Introduction

As empirical studies (e.g., Huang, 1992; Spink 1996) demonstrate, it is fairly common for information seekers to successively search over multiple sessions for essentially the same information problem. Choo, Detlor and Turnbull (1998) also found evidence of recurrent search episodes in studying Web search behavior of IT technical specialists, managers, researchers, marketing staffs, consultants, and firm administrative staffs. In an information system that provides a tremendous amount of information like the Web, failure to support successive searches likely results in frustration and inefficiency over the search process, as information seekers spend extra time and effort in recalling and re-searching for what were found and re-planning the continuous search process.

Theories help design interactive systems beyond intuitive judgments and with a more thorough understanding of users' task and problems (e.g., Shneiderman, 1998). However, theories that help justify the design space of system interface for supporting successive searches are still rare. For example, the GOMS model (Card, Moran and Newell, 1980, 1983) and its various follow-up models are useful in discussing the tasks of a specific search session, but can not explain how the users' goals could evolve over sessions. More importantly, they cannot account for how the evolving goal over sessions impacts on interaction with information systems and therefore cannot suggest useful support features.

In response to the void of theory in understanding successive search behavior, Lin and Belkin (2000) developed a conceptual model, called Multiple Information Seeking Episodes (MISE), identifying and characterizing eight modes of information problems as the reasons that prompt searchers to renew a search session.[1] These eight renewal reasons suggest eight different types of information tasks that could require different system supports.

This study focuses on one specific renewal reason and describes a prototyped information seeking toolkit that would support an information seeking task representative of the selected renewal reason. The particular selected renewal reason is called the

---

[1] Due to the space limitation, please refer to Lin (2001a) for a detailed version of MISE and Lin (2001b) for a concise version.

transmuting mode of successive searches, in which the searchers seek for information along the course of learning and clarifying their information problems.

According to the MISE model, for the transmuting mode of successive searches, the task of a searcher is to gather information for increasing the understanding of the problematic situation to plan for carrying out a task. Whenever a searcher cannot use his/her existing knowledge to accomplish the task, a problematic situation occurs. When a problematic situation becomes more urgent or salient, one has to attempt to cope with the deficient problematic cognitive state (PCS) by articulating information problems, evaluating search outcomes, and monitoring what constitutes such a PCS. When an searcher is thwarted from proceeding to one or more of these three processes by internal or external factors (e.g., need time to digest the information gathered, need to refer to books or people to grasp some specific aspects of the goal better), he/she stops the session and reinitiates the search when time is available.

Based on the conceptual system requirements derived from MISE, Personalized Indexing (PI) is prototyped as a tool enhancing successive searches over sessions for the transmuting information problems. This paper describes the task analysis, design requirements, system features, and an evaluation of PI and concludes with the possible directions to improve PI.

Task Analysis

The technique of cognitive task analysis (Roth and Woods, 1988) is utilized to derive from MISE the design requirements of PI for supporting the transmuting mode of successive searches. Cognitive task analysis is constituted by identifying (a) tasks the users perform, (b) problems the users encounter and (c) the support required to reduce those problems.

The tasks the user performs to search for information can be equated with those in the third dimension of the MISE model, namely, articulation, evaluation and monitoring (Marchionini, 1995), because these three sub-processes are the locus where the users actually physically interact with the system and where the system can actually intervene the users' information seeking activity.

Articulation of information problem can be generally carried out in three approaches: querying, browsing and "stepping stone." Querying is to represent the information problem with words or phrases. Browsing is to opportunistically tackle the information problem based upon information encountered. The "stepping stone" approach is to articulate information problem upon information attainment of previous episodes. For example, one can learn words from information attainment to help refine the information problem or follow links within information attainment to find other objects.

The sub-tasks of evaluation of search outcomes can entail developing relevance criteria, applying relevance criteria, and offline cultivating relevance criteria. Development of relevance criteria is to recognize aspects of the information problem to tackle; each aspect of the information problem can be a concept or topic that needs to be understood or explored. Application of relevance criteria is to examine the extent to which information objects contain information pertaining to concerned relevance criteria. Offline cultivation of relevance is to delay the development or application of relevance criteria until offline. The users can further evaluate collected information objects offline to either more carefully comprehend collected information or resort to other information resources, such as a friend.

Monitoring of problematic cognitive state can include three sub-tasks: estimating the search progress, making sense of search results, and continuous knowledge construction and concept development. Estimating the search progress is to examine what one has accomplished toward resolving the problematic situation. Making sense of search results is to know what one has done in order to continue the explication of the problematic situation. Continuous knowledge construction and concept development is to amend the problematic cognitive state and to coordinate processes of articulation and evaluation to determine when to stop and reinitiate a search episode.

The users could encounter a variety of problems within these three sub-tasks of information seeking that can be characterized in terms of properties of the four-dimensional MISE model (Lin, 2001). For example, searchers who successively search for transmuting information problems typically have a lower level of subject knowledge in the problematic situation dimension, are indefinite about their information problems in the information problem dimension, do not have clear criteria for evaluating usefulness of information objects, have to spend significant cognitive efforts in monitoring the status of information problem in the information seeking process dimension, and finally are prepared to resume the search in the episode dimension.

The rationale for the proposed support mechanisms is directly aimed to reduce the problems the users could encounter in three sub-processes of information seeking: articulation, evaluation and monitoring.

Design Requirements

The MISE model postulated that the prior knowledge state of the users would affect their current search experience. Moreover, search history of the previous episodes (abr. search history thereafter in this paper) consists of anything informative in assisting searches, ranging from used query terms, to information objects seen, to externally encoding objects (e.g., notes). Search history could be regarded as fair indication of the prior knowledge state of the users regarding their information problems. With MISE characterizing the transmuting mode of successive searches, the following high-level design requirements are inspired for an information system to utilize search history to support successive searches over multiple episodes.

- Search history as tangible entities:
  Search history needs to be tangible, so that one can see and even interact with it. For the transmuting mode of successive searches, a type of potentially useful object is explicit representation of relevance criteria, especially topicality. As the users do not exactly know what they are searching for, explicit and tangible representation of topicality as relevance criteria can help the users to keep control the progress of their information seeking.

- Search history as intelligence resources:
  Search history should be information assets potentially useful to improve the performance of the current and future searches. This reflects the underlying postulation in this study, that the knowledge of a searcher is constructive and accumulative across sessions; assimilation of knowledge contained in information objects seen in previous episodes would affect how a searcher defines his/ her information problems and undergoes the search.

- Search history as a persistent memory:
  Search history is like a super memory that never fades away. As the users in the transmuting mode of successive searches are in a learning process, they cannot remember a lot of things well, including where they have been, what they need to do, what they have found. Search history should be able to provide all the necessary answers.

- Search history as traceable footprints:
  Search history implies the track of traveling. Useful search history can allow one to return where he/ she visited before or to get an overview of what he/ she has done. Thus, easy historical navigation of a personal information space should be a system requirement for tools supporting the transmuting mode of searches (Wexelblat, 1999).

- Search history as customization craft:
  Search history, to some extent, is a product of selective memory. The users only remember what is relevant or what they want to remember. Therefore, the users should be able to manage and preserve selected portions of Search history.

- Search history as a fragment of the past:
  The users can only remember parts of the past rather than the whole. Thus, Search history should be flexible enough that the users can use multiple methods or incomplete piece of information to access Search history.

- Search history as a fragment of the current:
  Search history could be useful to the current, but only partially relevant because they have not kept updated with how the users' stock of knowledge has been developed since. Therefore, the system needs to provide methods to reduce the gap between the static knowledge contained in information objects and the dynamic knowledge states of the users that change over time.

- Search history as organizational portal:
  The users tend to group or associate information objects that connote the same knowledge frames together for the sake of sense making and knowledge construction (Baldonado and Winograd, 1997). This is especially true for the transmuting mode of successive searches since knowledge construction is one of the most significant characteristics of this type of search behavior. Moreover, the users would organize objects to help remember how to navigate their personal information space.

- Search history as a learning system/ wisdom mine:
  People ponder over Search history for lessons. They want to gain a sense of what has been accomplished. The interface needs to learn more about the users than the users themselves. It would not only help the users to see what they should have seen, but also communicate with the users about the search process. System feedback apparently is necessary.

- Search history as a gesture for the future:
  If an information object encountered seems useful, the users tend to mark it in some ways as some sorts of memory aids so that when encountering it later, they can quickly pick up from where they left off (Marshall, 1997). In the transmuting mode of successive searches, the users are likely to have ill-defined information problems and a lower level of domain knowledge. Revisiting information objects for follow-up relevance examination is important, and so is knowing how to continue the search. Thus, it is important to support the users not only in utilizing Search history, but also in preparing for future searches. Annotation-like objects can be a potential solution for connecting the current search to future ones.

Design Product

With the high-level design requirements unveiled above, features of PI are specified. The fit between the design requirements and the tool is described in this section by discussing functionalities, problems improved, and design rationale.

With the design requirement of search history as tangible entities, PI is designed as the indexing mechanism that allows the users to explicitly and tangibly develop their knowledge network hierarchy in terms of their information problems; each item in such a knowledge network hierarchy is a concept or topic representing an aspect or facet of information problems. The ideas behind it are to support the users in developing and applying indefinite relevance criteria over multiple search sessions and in estimating and making sense of the information attainment subordinated to those criteria. It works in a way in which the users are indexing the information objects relevant to their information problems and at the same time transparently storing and organizing relevant information objects, creating a meaningful personal information space.

With the design requirement of search history as a gesture or invitation for future, PI is designed to function like informal coding of information objects, which is a form of annotation for the purposes of place marking and aiding memory (Marshall, 1997). PI supports place marking by allowing the users to mark relevant information objects with indexing terms, where an indexing term represents a relevant concept tackling the information problem. It supports aiding memory with index terms, which reminds of the users what aspects of information problems they have tackled and what their information attainment is about. In addition, all the information objects would be grouped with the same index terms. It would be easy for the users to assess how useful their information attainment for different aspects of their information problems is.

With the design requirement of search history as traceable footprints, the system needs to provide a mechanism to support the users following footprints back to any point (i.e., information object seen) in the past episodes. The index terms in PI are the footprints the users left in the past episodes, indicating certain information objects as potentially useful and reminding the users of what information objects are about. PI would allow the users click at those indexed information objects and return to them.

With the design requirement of search history as a fragment of the past, the system needs to enable the users with multiple options of returning to information objects visited. In addition to tracing indexing terms, the system would enable the user to use key concepts as query terms to search against the pool of information objects seen. The key concepts are segment of Search history, may or may be consistent with index terms, but the system would fully index information objects to enable the search option in locating information objects seen.

The design requirement of search history as organizational portals leads to the function of organizing a personal information space in PI. PI would enable the users to dynamically group information attainment in terms of their information problems at hand, in order to easily evaluate the topicality strength of information attainment and quickly locate relevant information objects by visualizing their personal information space.

Knowing the topicality strength of information attainment and assimilating information from relevant information objects, the users could better reformulate their information problems and construct knowledge. Furnas and Rauch (1998) noted, "Organizational activities allowed users both to clarify their evolving conceptualization and to coordinate their search activity." By organizing information attainment, the users increase their comprehension and assimilation of information attainment because they are driven to examine their information attainment for their transmuting information problems. The assimilated information

further changes the users' stock of knowledge, increases the understanding of their problematic situation, and assists in assessing their information problems.

Thus, the annotation and organization functions, combined together, correspond to the design implication of Search history as a learning system or wisdom mine. They help the users to determine how much they have achieved so that they can decide how to continue to explore the information space.

An implicit function of PI is storage of information, giving rise to the design requirement of search history as persistent memory. What are stored are index terms, locators of information objects, and other attributes of access to information objects (e.g., access time). The storage function enables the system to display the relationship of indexing terms with information attainment. Thus, the storage function is presumed necessary to the organization function.

As the users are organizing their indexing scheme, they also likely need to re-organize information objects subordinated to the indexing scheme, as the design requirement of Search history as customization craft indicates. For example, after the modification of the indexing scheme, an information object originally subordinated to an indexing term A may be more suitable for another indexing term B. Basic functions, such as moving and deleting, would be necessary for organizing information objects.[2]

With the design requirement of search history as a fragment of the current episode, the system needs to help the users to update their indexing scheme because information attainment evaluated might not have been evaluated with the criteria used in the current episode and updated with the growing subject knowledge in representing information objects. The system could provide an "update" function. When the users invoke it, the system may suggest terms for users to index those non-indexed information attainments and update index terms for indexed information attainment by computing the similarity of information objects that were assigned with certain indexed terms and those information objects which have not.[3]

Table 1 summarizes how the requirements of search history correspond to the MISE model and help suggest the proposed system features, which are integrated as Personalized Indexing.

Table 1. How Functionalities of PI Are Derived

Requirements of Search history	Properties of MISE	System Features
Tangible entities	Relevance criteria	User Indexing
Intelligence resources	Stock of Knowledge	Annotation
Persistent memory	Ill-sedimentation as a problematic cognitive state	Storage of search history
Traceable footprints	Constructive knowledge	Indexing as annotation
Customization craft	Relevance, topicality	Dynamic re-organization of information
A fragment of the past	Remembrance as a problematic cognitive state	Browsing and searching information objects seen
A fragment of the current	Gaps between information objects and evolving knowledge state	Updating indexing terms
Organizational portal	Topicality of knowledge; knowledge construction	Information organization
A learning system/ wisdom mine	Constructive knowledge	Information organization and annotation
A gesture for the future	Constructive knowledge	Annotation

---

[2]In this version of PI, the moving and deleting functions are not prototyped, due to the limitation of the development time.

[3]In this version of PI, the "update" function is not prototyped, due to the limitation of the development time.

Use Scenario

A typical scenario of successive searches for transmuting information problems is to search for information in order to plan a vacation and decide the final destination and itinerary among a few destinations that a searcher knows little about. It is transmuting because the searcher has little subject knowledge to start the search and has to learn about what to search during the course of information seeking and base the search decision largely on previous information attainment. Below is an example of how John uses PI to enhance his successive searches for a transmuting vacation planning task.

When John sees a Web page that he thinks useful to explicate his information problem or that he might revisit to further examine the content or follow links in the near future, he would type in the terms (e.g., New York city, guide) representing the Web page in regard to his information problem in the text entry box to index the page, as figure 1 shows. John can continue to index all the useful Web pages in this manner. The Web pages indexed with the same index terms will be grouped together under those index terms, as shown in figure 2.

As the information attainment grows, the tree-structured representation of information attainment could become too large to view in a quick glance. PI provides four organizational methods to display a better overview of the information attainment as needed. First, when John shifts the focus of his information problem from one destination to another, he could close all the branches of non-focused destinations (e.g., New York and Seattle) and only open branches of the focused destination (e.g., Orlando).

Second, John could use the "group by" function to group information attainment in terms of destinations so that he could see the strength of his information attainment for each destination by entering "New York City, Seattle, Orlando" in the pop-up window invoked by the "group by" function. As shown in figure 3, the index terms other than grouped-by terms form another level of headings. For example, all other index terms associated with New York City will be grouped under "New York City." This way, John can easily see different types of information found for different cities. Also, John can see that his information attainment about Seattle is more complete (e.g., more options for lodging) and detailed (e.g., various types of dining) than other cities. John can also change to organize his information attainment by activity (e.g., lodging, night life, dining) in the first level and by destination in the second level, simply performing the "group by" function twice, as shown in figure 4.

The third organization method, "index one," lists all index terms one at time and all in the same level. Web pages that are indexed with multiple index terms will appear multiple times under those separate index terms. Figure 5 shows an example of "index one," in which John can see how many Web pages he has collected for each destination, regardless of the types of activity, or how many Web pages for each activity, regardless of the destination. Such an organization method could facilitate comparison of related Web pages, as all the same activity or destination information will be grouped together.

Opposed to "index one" is the "index all" function, which enumerates all index terms associated with the same Web pages together as the way John indexes them. Figure 6 shows an example of the "index all" result. This organization method enables John to find a particular Web page more easily, as the full index terms could remind the user of what each Web page is about.

Experimental Results

A controlled experiment was conducted for two reasons. First, it helped evaluate the validity of system requirements derived from the MISE model. If as predicted in MISE, the subjects sought for a great deal of support in browsing, evaluation and monitoring, then MISE as a framework of task analysis for analyzing the transmuting mode of successive searches indeed is validated. In addition, the experiment explored reasons why subjects used PI, which would help understand what are important in successive searches, indirectly verifying the MISE model. Second, the experiment also compared PI to existing similar system features. The comparison results would help understand how to improve PI.

20 college students were recruited to participate in the experiment, consisting of three search sessions, each of which lasted between 20 and 30 minutes. Before the experiment, subjects received a tutorial that explains the functions of PI. At the beginning of each session, subjects were given a description of session tasks and goals. Subjects were interviewed at the end of the experiment with 7-point Likert-Scale questions to collect data about the extent to which PI was helpful in coming up with query terms, navigating the information space, estimating and projecting the search progress, and organizing search thoughts. For mental model of PI, subjects were asked to identify what they intended to do with PI.

Table 2. Search Performance with PI

	Querying	**Navigating**	**Estimating**	**Organizing**
Mean	2.33	6.11	5.53	5.76
Std Div	1.68	.74	1.65	.97

Confirming the prediction of MISE, the results, as in table 1, show that subjects generally received a great deal of support from PI in navigation, estimation/ projection, and organization, but less in representation of information problems. In addition, 15 subjects indicated that they used PI primarily for revisiting Web page, a type of navigation activity, 10 for visualizing the information task for the mixed purposes of estimation of search progress and organization of search thoughts, and 5 for organizing information collection, a type of organization activity. Together, the results suggest that PI successfully supports subjects in the activities that MISE describes as important in successive searches.

For the comparison question, subjects were asked to compare PI to Bookmarks of Netscape and Favorites of Internet Explorer and specify their preference. Bookmarks/ Favorites were compared because like PI, they provided capabilities of navigation and organization (Abrams, Baecker, and Chignell, 1998). 12 subjects thought that PI was significantly different from Bookmarks/ Favorites for the reasons of better organizational functions, visibility, and ease of finding Web pages to return. 3 subjects were neutral in judging the difference. 5 subjects could not rank the difference mainly because they had little experience in Bookmarks/ Favorites, although 1 did not use PI at all. The results appear to indicate that PI is superior over other similar features for its organization and navigation capabilities.

The nature of the study is exploratory rather than explanatory. All the findings have their limitations in generalization. For example, the number of participating subjects is small and homogeneous, only 20 and all college students majoring in communication. The subjects were studied under a controlled experiment, videotaped and asked to think aloud. The search atmosphere could have been intrusive and affected subjects' search performance.

Conclusions and Future Work

This study has prototyped PI to support the transmuting mode of successive searches, based on the requirements derived from the MISE model. The preliminary empirical results validate MISE as it found that PI is successful in supporting navigating the information space and monitoring the problematic cognitive state, two important search sub-tasks for transmuting information problems. The author is encouraged to continue development of PI with additional functionalities such as deleting and updating coding schemes in PI.

Continuous studies are also planned to explore the relationship between performance and productivity of successive searches, for example, whether the support in navigation and monitoring can contribute to efficiency and effectiveness of searches, why or why not. If not, what needs to be done to transform an enhanced search experience to an improved search consequence. The results of this study and continuous evaluation studies in plan will help revise the requirements of PERSIST and improving the design space for a tool that supports successive searches across multiple sessions.

Acknowledgements

References

Abrams, D., Baecker, R.M., and Chignell, M. Information Archiving with Bookmarks: Personal Web Space Construction and Organization, in *Proceedings of the CHI '98 Conference,* ACM Press, 1998, pp. 41-48.

Baldonado, M. Q. W. and Winograd, T. Sense maker: An information-exploration interface supporting the contextual evolution of a user's interests, in *Proceedings of the CHI '97 Conference* Atlanta GA, ACM Press, 1997.

Barrett, R. and Maglio, P. P. Intermediaries: New places for producing and manipulating web content. *Proceedings of the Seventh International World Wide Web Conference (WWW7)*, Brisbane Australia, 1998.

Barrett, R., Maglio, P. P., and Kellem, D. C. How to personalize the Web, in *Proceedings of the CHI '97 Conference*, Atlanta GA, ACM Press, 1997.

Card, S., Moran, T. P. & Newell, A. The keystroke-level model for user performance with interactive systems. Communication of the ACM, 23, 1980, pp. 396-410.

Card, S., Moran, T. P. & Newell, A. The psychology of human computer interaction, Lawrence Erlbaum Associates: Hillsdale, NJ, 1983.

Choo, C. W., Detlor, B. & Turnbull, D. A Behavioral Model of Information Seeking on the Web- Preliminary Results of a Study of How Managers and IT Specialists Use the Web. *Proceeding of the 61st ASIS Annual Meeting*. Pittsburgh PA. 1998, pp. 290-302.

Furnas, G. W., and Rauch, S. J. Considerations for information environments and the NaviQue workplace. *Digital Library 98*, Pittsburgh PA, 1998, pp. 79-88.

Huang, M. H. Pausing behavior of end-users in online searching, Unpublished doctoral dissertation. University of Maryland. 1992.

Lin, S. J. and Belkin, N. J. Modeling Multiple Information Seeking Episodes, *Proceeding of the 63rd ASIS Annual Meeting,* Chicago, IL, 2000, pp. 133-147.

Lin, S. Modeling and supporting multiple information seeking episodes over the Web, Unpublished doctoral dissertation, Rutgers University. 2001a.

Lin, S. J. Understanding Successive Searches across Multiple Sessions over the Web, In *Proceedings of the Twenty-Second International Conference on Information Systems*. New Orleans, LA, 2001b.

Marchionini, G. *Information seeking in electronic environments*. Cambridge University Press, New York NY. 1995.

Marshall, C. Annotation: from paper books to the digital library. *Proceedings of the ACM Digital Libraries '97 Conference*, Philadelphia, PA. 1997.

Nielsen, J. Why frames suck (most of the time), Available: http://www.useit.com/alertbox/9612.html. 1996.

Roth, E. M., and Woods, D. Cognitive task analysis: An approach to knowledge acquisition for intelligent system design. In G. Guida & C. Tasso (ed.). *Topics in expert system design: Methodology and tools,* North-Holland, Amsterdam, 1988, pp. 233-264.

Shneiderman, B. Designing the user interface: Strategies for effective human-computer interaction. (3rd Ed.) Addison Wesley Longman, Inc. 1998.

Spink, A. Multiple search sessions model of end-user behavior: An exploratory study. *Journal of the American Society for Information Science*. (47: 8), 1996, pp. 603-609.

Wexelblat, A., and Maes, P. Footprints: History-Rich Tools for Information Foraging, in *Proceedings of the CHI '99 Conference* (Pittsburgh, PA), ACM Press. 1999.
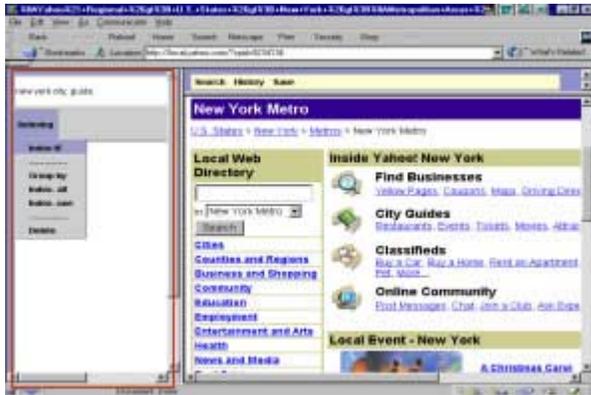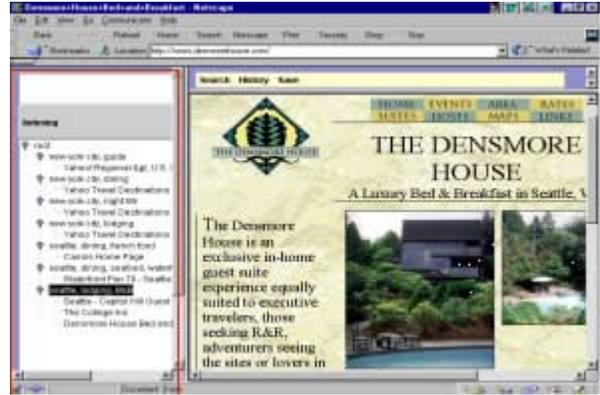
Figure 1. Screen Shot of Using the "Index It" Function

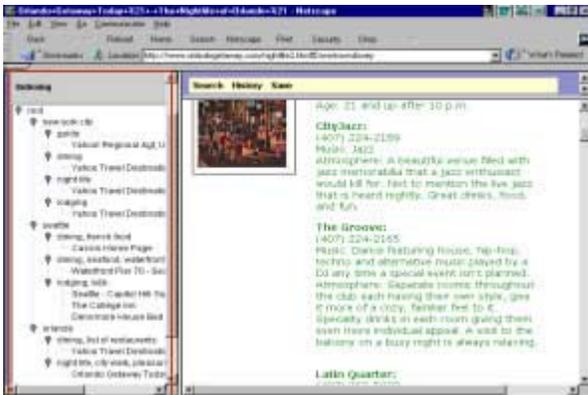**Figure 2. Screen Shot of Multiple Web Pages Grouped under the Same Index Terms**

Figure 3. Screen Shot of Result of the "Group By" Function

Figure 4. Screen Shot of Result of Multi-level "Group By"

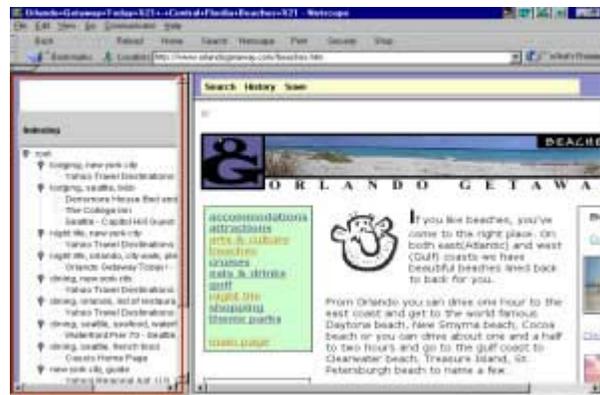**Figure 5. Screen Shot of Result of the "Index One" Function**

Figure 6. Screen Shot of Result of the "Index All" Function