

Content and Knowledge Management in a Digital Library and Museum

Jian-Hua Yeh, Jia-Yang Chang, Yen-Jen Oyang

Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, Republic of China. E-mail: yjoyang@csie.ntu.edu.tw

This article discusses the design of a digital library that addresses both content and knowledge management. The design of the digital library features two major distinctions: (1) the system incorporates a two-tier repository system to facilitate content management, and (2) the system incorporates an object-oriented model to facilitate the management of temporal information and exploits information extraction and deductive inference to derive implied knowledge based on the content of the digital library. The two-tier repository system relieves the system manager from manually maintaining the hyperlinks among the Web pages, when the digital library content is updated. The task of maintaining hyperlinks among Web pages can become cumbersome to the system manager if there are a large number of Web pages and hyperlinks. With respect to knowledge management, this design aims at facilitating temporal information management and deriving implied relations among the objects in the digital library. The motivation behind developing these knowledge processing utilities is to create a system that complements the capabilities of human beings. Deriving a comprehensive list of implied relations is an exhausting task if the digital library contains a great amount of information and the number of implied relations is great. With such knowledge-processing utilities, specialists are released from performing tedious work and can, therefore, spend more time with more productive philosophical activities to derive advanced knowledge. Applying knowledge management utilities effectively can extend the applications of digital libraries to new dimensions.

Introduction

Advances in computer network and storage technologies have inspired the design of digital libraries in recent years (Arms, Blanchi, & Overly, 1997; Gladney et al., 1994; Kahn & Wilensky, 1995; Nuernberg, Furuta, Legget, Marshall, & Shipman, 1995; Schatz & Chen, 1996). The emergence of digital libraries has introduced a number of important issues (Daniel & Lagoze, 1997; Glymour, Madigan, Pregibon, & Smyth, 1996; Imielinski & Mannila, 1996). One issue that

has not attracted much attention, but is essential to digital library development, is the management of digital library contents. Because a typical digital library contains a large quantity of documents, the conventional practice used to maintain the contents of web sites will not work for digital libraries. In particular, the hyperlink that is widely incorporated in conventional web sites to structure the web pages poses a major challenge. For a digital library with a large quantity of materials, attempting to manually maintain the hyperlinks among web pages will become a cumbersome task for the manager of a digital library. Due to this observation, the design of the National Taiwan University Digital Library and Museum (NTUDLM) incorporates a two-tier repository architecture that eliminates the need to manually maintain hyperlinks among the object Web pages.

Another interesting issue that emerged with the development of digital libraries is knowledge management. In some cases, if the digital library can also provide knowledge management capabilities (Chen, 1990; Chen, Smith, Larsgaard, Hill, & Ramsey, 1997; Wong et al., 1997), then the application of the digital library is extended to a new dimension. The design of the NTUDLM aims at facilitating the management of temporal information and exploiting information extraction (Cowie & Lehnert, 1996; Solderland, 1996; Solderland, Fisher, & Lehnert, 1997) and deductive inference (Fagin, Halpern, Moses, & Vardi, 1995; Russel & Norvig, 1995) to derive implied knowledge based on the content of the digital library. One application of the incorporated knowledge-processing utilities is to derive the implied relationships among the objects in the digital library from a large volume of information. The following is an example. Assume that one piece of information in the digital library reveals that a person named A served in a government agency for some time, and another piece of information reveals that a person named B also served in the same agency for some time. Should the periods during which A and B served in the agency overlap, then these two pieces of information together imply that A and B were colleagues for a period of time. Deducing a piece of this kind of information may be quite trivial to most people. However, if the

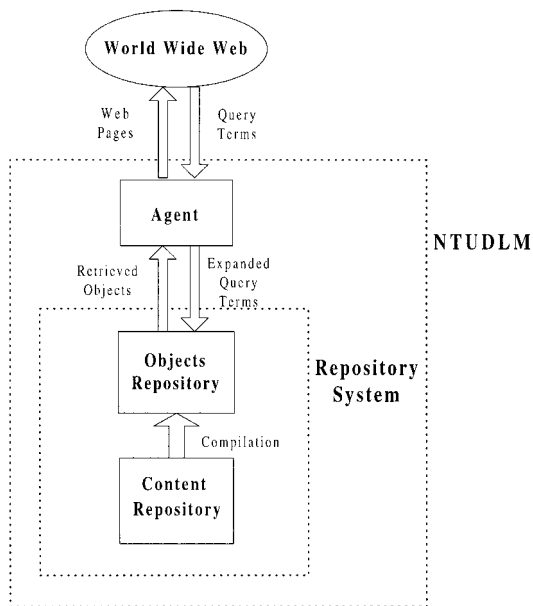


FIG. 1. System architecture of the NTUDLM.

digital library contains a great amount of information and the implied relationships are extensive, completing this kind of task may exhaust the person who attempts to do it manually. With the incorporation of knowledge-management utilities, specialists are released from carrying out such tedious work, and can spend more time in the pursuit of philosophical activities to derive more advanced knowledge.

To facilitate the management of temporal information and derive implied knowledge, an object-oriented model was developed. In recent years, there have been several articles that address the representation of temporal information using object-oriented structures (Goralwalla, Özsu, & Szafron, 1997; Goralwalla, Leontiev, Özsu, & Szafron, 1997; Tong, Wüthrich, & Sankaran, 1995). However, the developments in those structures aimed at the recording and retrieval of temporal information in databases. Because the main issue addressed in this article is deriving implied knowledge based on temporal information, an object structure was developed with this goal in mind. In the process for deriving implied knowledge, the temporal logic system developed by James F. Allen was employed (Allen, 1991).

In the following discussion, the next section elaborates on the structure of the two-tier repository system and the content management mechanism employed at NTUDLM. Then the following section describes the object-oriented model developed to facilitate management of temporal knowledge. The Information Extraction Section addresses how information can be automatically extracted from the content of a digital library and the process for deriving implied knowledge based on temporal information. Application of the Knowledge Processing Utilities provides examples of knowledge management utility applications. The System Implementation discusses system implementation

and evaluation. Finally, concluding remarks are presented in the Conclusion Section.

Structure of the Two-Tier Repository System

Figure 1 shows the major components of the NTUDLM. There are three main modules: the agent, the objects repository, and the content repository. The agent is the module that directly interacts with the user. Upon receiving a query from the user, the agent expands the query based on the thesaurus (Chen, Yim, Fye, & Schatz, 1994; Miller, 1997; Park & Choi, 1997; Schatz, Johnson, Cochrane, & Chen, 1996; Schutze & Pedersen, 1997) of the NTUDLM content and then sends the expanded query to the repository system. Figure 2 exemplifies the query expansion carried out by the agent. In this example, Zealandia, Tayouan, and Tayovan are ancient names of a place now called An-Ping. Because a user may not know all of An-Ping's ancient names, by which An-Ping is referred to in ancient documents, carrying out a query expansion with the thesaurus is an effective way to improve the recall rate of the search result. The repository system will retrieve all related materials based on the expanded query and return the retrieved materials to the agent. The agent will then integrate the retrieved materials to generate a Web page.

One of the main distinctions in the NTUDLM system design is that a two-tier repository system is incorporated. The two-tier repository system comprises the objects repository and the content repository. The main reason why the two-tier design is employed is to facilitate content management. The objects repository stores and manages the content the NTUDLM in an object-oriented form, and is responsible for resolving information retrieval requests on the fly. To the agent, the objects repository is an object-oriented database.

The content repository stores and manages the content of the NTUDLM in a relational database. The content repository does not respond to information retrieval requests on the fly, which is the objects repository's responsibility. The main role that the content repository plays is to facilitate content management. When updates to the NTUDLM content are to be carried out, the updates are first made to the content repository. Then, the content of the content repository

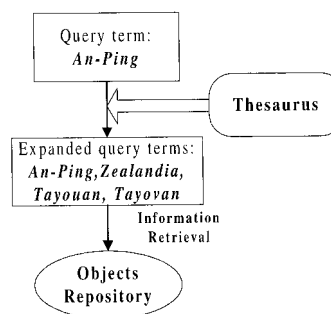


FIG. 2. Query expansion example.

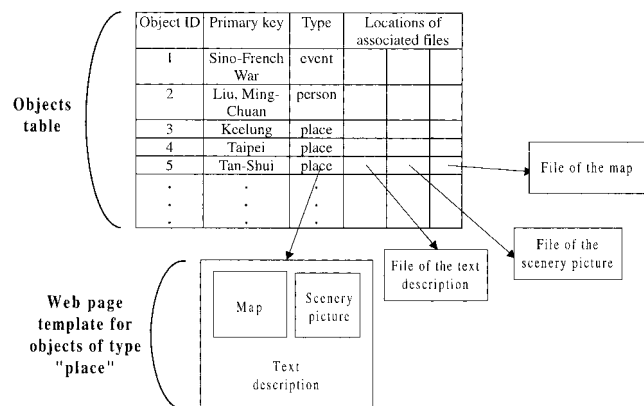


FIG. 3. The structure of the content repository.

itory is compiled to generate a new object structure for the objects repository. During the generation process, crossreferences among the descriptive data of the objects are established. This practice eliminates the need to manually maintain hyperlinks or crossreferences among the descriptive Web pages of the objects.

Figure 3 depicts how the objects in the NTUDLM is stored in a relational database in the content repository. Each object in the NTUDLM has an entry in the objects table. The objects are classified based on the types of Web page templates that will be used to synthesize the Web pages of the objects. The associated files fields of each object specify the files that will be invoked in synthesizing the Web page of the object. The table shown in Figure 3 is maintained by the manager of the content repository. When a new object is to be added into the content repository, the manager needs to fill out the fields in the table.

What are missing in the structure shown in Figure 3 are the crossreferences among the Web pages contents of the objects. In the NTUDLM, the crossreferences are established using an annotation utility as shown in Figure 4. The annotation utility examines the text files whose contents are to be incorporated on Web pages and looks for instances of the primary key terms in the table shown in Figure 3. At each instance of the primary key terms, the annotation utility annotates a CGI (Common Gateway Interface) tag.

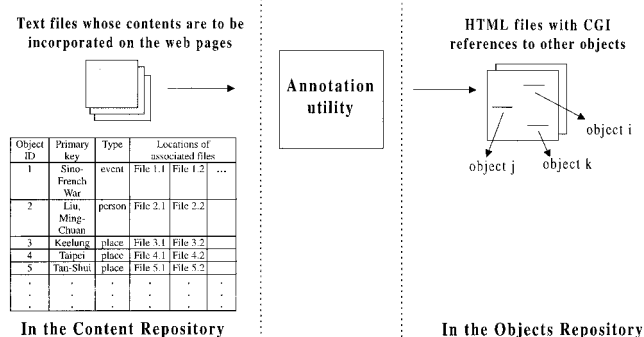


FIG. 4. Crossreferences established by the annotation utility.

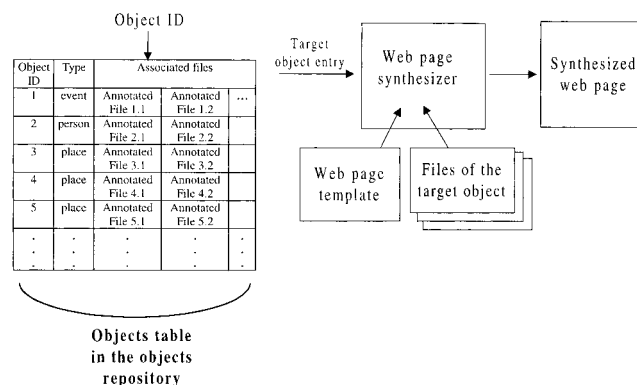


FIG. 5. Process of synthesizing a Web page by the objects repository.

When a CGI tag on a Web page is clicked, a query containing the primary key of the tag is sent to the objects repository.

The HTML files with CGI annotations, which are generated in the annotation process shown in Figure 4, are stored in the objects repository. When an object is invoked by the agent, the objects repository retrieves the corresponding Web page template according to the object type and the files associated with the object to synthesize the Web page of the object. Figure 5 illustrates the process of synthesizing the Web page of an object on the fly. During the synthesis process, the text files used are those that have been annotated with CGI tags, not those plain text files stored in the content repository.

The main advantage that the practice described here offers is a very effective method for managing the contents of a digital library. When the manager of the digital library wants to add new objects or modify the descriptive texts for some objects, the manager only needs to make updates to the content repository. Then, the annotation utility will automatically establish all crossreferences among the descriptive texts for the objects. This eliminates the cumbersome task of manually maintaining the hyperlinks or crossreferences among Web pages.

Organization of Temporal Knowledge in the NTUDLM

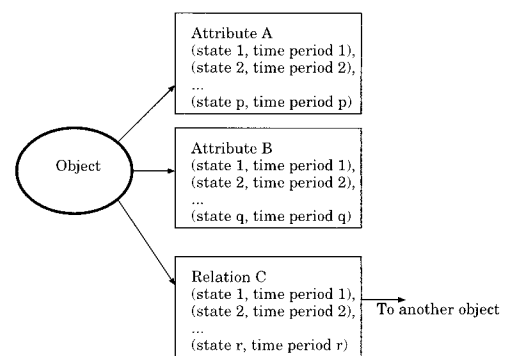
The discussion in the last section focuses on how the contents of the NTUDLM are structured to support browsing operations conducted by the users. This section will address the object-oriented model employed in the objects repository to facilitate the management of temporal knowledge. Because of the nature of the NTUDLM, which contains mainly historical materials, temporal knowledge is the most important type of knowledge in this application. One example of temporal knowledge is the relationship between two people. Two people may have been classmates when at a school, and subsequently, colleagues, during their career development. As a result, time qualifiers are needed to precisely describe the classmate and colleague relationships

between these two persons. In the past, there have been a lot of studies on the representation and reasoning for temporal knowledge (Allen, 1991; Galton, 1987; Goralwalla et al., 1997a; Vila, 1994). What is addressed in this article is an object-oriented model developed to facilitate the management of temporal knowledge.

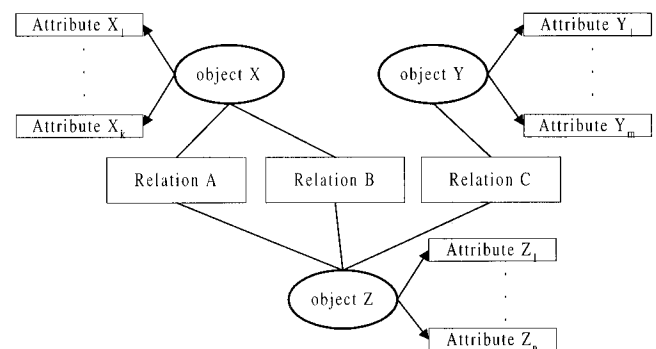
In the object-oriented model of the NTUDLM objects repository, there are two categories of objects, namely the informative object, and the cognitive object. Informative objects are the building primitives of the NTUDLM. Typical informative objects include paragraphs of text, pictures, maps, video programs, etc. Each informative object has its own well-defined contents, and the contents normally do not change over time. Informative objects collectively are not structured except that they may be classified according to their nature. However, classification of informative objects is simply for easy management, and has no structural implication.

Informative objects define information piece by piece. However, in human cognition, there is a higher layer of abstraction. In the NTUDLM, objects of the higher layer of abstraction are called cognitive objects. Typical cognitive objects include persons, events, places, etc. A cognitive object is a concept described by a collection of related informative objects, and a cognitive object itself does not have well-defined contents. Human's interpretation of a cognitive object is more dynamic than that of an informative object. As mentioned earlier, each informative object has its own well-defined contents, which normally do not change over time. On the other hand, the meaning and characteristics of a cognitive object may change from time to time. When new information related to a cognitive object is found and added to the digital library, how the cognitive object stands in our interpretation may change.

A typical informative object or cognitive object is associated with one or more attributes and relationships. The attributes and relationships are the structures used to store temporal information in the NTUDLM. The attributes associated with an object describe the object from various aspects. Each attribute of an object is described by a sequence of state tuples. A state tuple contains a state and a period of time during which the state is valid. The state tuples in a sequence are sorted according their order in time domain. Some attributes are valid indefinitely. Such attributes are described by single-state sequences in which the states are valid indefinitely. While an attribute describes an object from one aspect, a relation associates one object with another object. A relation can exist between two informative objects, between two cognitive objects, or between one informative object and one cognitive object. Because two objects in the real world may be related in different aspects, more than one relationship can coexist between two objects. For example, two persons can be both relatives in their family life and colleagues in their career development. In such situations, it is better to have two relations to describe these two persons' relationship from the two different aspects. A relationship, just like an attribute, is described by a



(a) The structures of a typical object.



(b) The structures among objects.

FIG. 6. Object-oriented model in the NTUDLM.

sequence of state tuples. Figure 6(a) depicts the attribute and relation structures of a typical object. Figure 6(b) illustrates the structures among the objects in the NTUDLM.

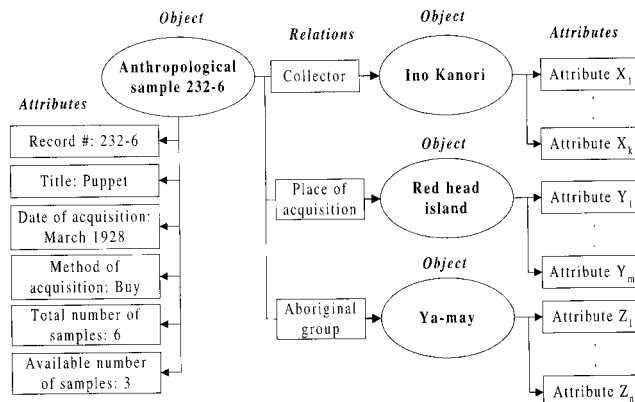
Like the crossreferences among the Web pages, the attribute and relation structures of the objects are derived from compiling the content in the content repository. In the content repository, there are one attributes and one relations table that record all the attributes and relations, respectively.

Information Extraction and the Process of Deriving Implied Knowledge

The object structures presented in last section were developed to facilitate the management of temporal knowledge. The next question is how the desired temporal knowledge is acquired. The NTUDLM relies upon two sources to acquire knowledge. The first source is from the metadata (Baldonado, Chang, Gravano, & Paepcke, 1997; Weibel, 1995) associated with the objects in the digital library. The second source is through information extraction. In the NTUDLM convention, each object is associated with a set of metadata. The metadata is created manually by contents specialists and contains knowledge that can not be easily derived by information extraction utilities. Figure 7(a) shows an example of metadata. Some metadata fields spec-

Record #	232-6
Title	Puppet
Aboriginal group	Ya-may
Collector	Ino Kanori
Date of acquisition	March 1928
Place of acquisition	Red head island
Method of acquisition	Buy
Total number of samples	6
Available number of samples	3
.	.
.	.

(a) An example of metadata.



(b) The objects structure based on the metadata in (a).

FIG. 7. An example of metadata from NTUDLM.

ify the attributes of the object and some specify how this object is related to other objects. Figure 7(b) depicts how the knowledge presented in the metadata in Figure 7(a) is recorded in the object structure described in the last section.

Figure 8 shows an example from a history book that is a good case to which to apply the information extraction utility. The example summarizes the backgrounds and careers of the officers that served in the Taiwan county government during the Chin Dynasty. It is apparent that the list exhibits certain rules. For example, in the description of each officer, the first time period always specifies when the officer took the job, and the second time period always specifies when the officer left the job. Therefore, it is quite straightforward to automatically extract information about the time intervals during which these officers served in their jobs.

As far as deriving implied knowledge is concerned, the temporal logic developed by James F. Allen (Allen, 1991) is employed. In the temporal logic, a time argument is added to the conventional predicate to form the temporal predicate of form $P(a_1, a_2, \dots, a_k; t)$, where a_1, a_2, \dots, a_k are nontemporal arguments, and t is the interval during which the predicate is of concern. Figure 9(a) illustrates various relations between two time intervals identified by James F. Allen, and Figure 9(b) shows the primitive operations between two intervals.

The deductive inference utility first employed in the NTUDLM project is for deriving the relations among the objects in the digital library that the temporal information

County governor
Shen, Chao-pin
Born in Liaotong. Assumed in 1644.
Jiang, Hsiang
Born in Liaotong. Also known as Ke-ming. Assumed in 1646 and terminated in 1648.
Wang, Chao-sheng
Born in Tong state in Jiangnan. Education background: Ju-ren. Assumed in 1648 and terminated in 1651.
Wang, Shi-june
Also known as Ju-an. Assumed in 1665.

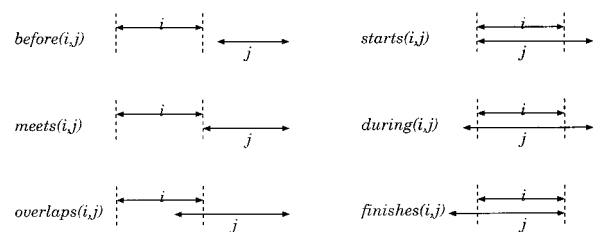
Secretary-general
Chao, Shing-ke
Born in Shanshi Province. Education background: Gong-shen. Assumed in 1644 and terminated in 1648.
Chen, Jia
Born in Chejiang Province. Also known as Mei-chuan. Education: Jian-shen. Assumed in 1649 and terminated in 1651.
Chang, Yuan-ying
Born in Liaotong. Assumed in 1651.
Wang, Li-chong
Also known as Shu-jai. Born in Yi county in Jiangnan. Assumed in 1665.

Executive secretary
Chang, Yuan-chu
Born in Chao County. Also known as Huan-an. Education background: Li-yuan. Assumed in 1644 and terminated in 1648.
Kao, Lang
Born in Shantung Province. Education background: Li-yuan. Assumed in 1649 and terminated in 1652.
Lee, Ting-quai
Also known as Min-jia. Born in Tong in Jiangnan. Assumed in 1665.

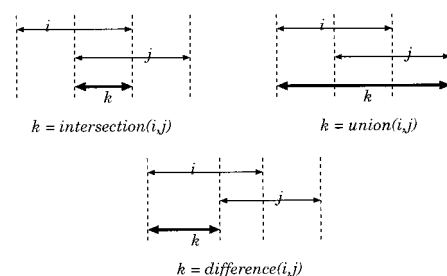
Director of police division
Chie, Weng-da
Born in Shanshi Province. Also known as Chien-ren. Education background: Li-yuan. Assumed in 1644 and terminated in 1648.
Chang, Weng-mo
Born in Shantung Province. Education background: Li-yuan. Assumed in 1649 and terminated in 1652.

FIG. 8. An example from a history book.

stored in the attributes and relations of objects implies. The process of deriving implied relations starts with a set of m objects, denoted by $\{O_1, O_2, \dots, O_m\}$, and a set of n



(a) The different interval relationships.



(b) Some primitive operations between two intervals.

FIG. 9. Temporal operations.

```

for every inference rule  $R_i \equiv P_i \Rightarrow Q_i$ 
{
     $S_1$ =set of objects that the first argument of  $P_i$  concerns;
     $S_2$ =set of objects that the second argument of  $P_i$ 
concerns;
    ...
     $S_k$ =set of objects that the  $k$ -th argument of  $P_i$ 
concerns;
    for every combination of objects  $(O_{i1}, O_{i2}, \dots, O_{ik})$ ,
    where
     $O_{i1} \in S_1, O_{i2} \in S_2, \dots, O_{ik} \in S_k$ 
    {
        for every combination of the states of
         $(O_{i1}, O_{i2}, \dots, O_{ik})$  that  $P_i$  concerns
        {
            evaluate  $P_i$ ;
            if the evaluation results in "True"
            {
                add the newly derived relation into
                the object structures;
            }
        }
    }
}
end procedure;
```

FIG. 10. Pseudocode for deriving implied relations among objects.

temporal inference rules, denoted by $\{R_1, R_2, \dots, R_n\}$. These m objects are structured in accordance with the object structures presented in the last section, and each of these n temporal inference rules specifies how a new relationship can be derived from the attributes and existing relations recorded in the object structures. Following is an example of the temporal inference rule. The example derives a colleague relationship between two persons based on their career histories. Figure 10 shows the pseudocode of the deductive inference process.

$$\text{Job}(\text{Person}_A, \text{Company}_X; T_1) \wedge \text{Job}(\text{Person}_B, \text{Company}_X; T_2) \wedge \text{overlaps}(T_1, T_2) \Rightarrow \text{Colleague}(\text{Person}_A, \text{Person}_B; \text{intersection}(T_1, T_2))$$

Applications of the Knowledge-Processing Utilities

The discussion in this section first uses a simple example to illustrate the process of deriving implied relationships

TABLE 1. Officers and their positions.

Name	Position	From (year/month)	To (year/month)
Shen, Chao-pin	County governor	1644/01	1644/01
Jiang, Hsiang	County governor	1646/01	1648/01
Wang, Chao-sheng	County governor	1648/01	1651/01
Wang, Shi-june	County governor	1665/01	1665/01
Chao, Shing-ke	Secretary-general	1644/01	1648/01
Chen, Jia	Secretary-general	1649/01	1651/01
Chang, Yuan-ying	Secretary-general	1651/01	1651/01
Wang, Li-chong	Secretary-general	1665/01	1665/01
Chang, Yuan-chu	Executive secretary	1644/11	1648/11
Kao, Lang	Executive secretary	1649/01	1651/01
Li, Ting-quai	Executive secretary	1665/01	1665/01
Chie, Weng-da	Director of police division	1644/01	1648/01
Chang, Wen-mo	Director of police division	1649/01	1652/01

The dates are in Lunar calendar.

from temporal information stored in a digital library. Then, a number of situations are presented to demonstrate how such knowledge processing utilities can be applied to extend the applications of digital libraries.

The example used is to figure out the bureaucratic relationship among the government officers whose career records are present in Table 1. Figure 11 shows the bureaucratic hierarchy of the county government of Taiwan during the Chin dynasty. In this example, the criterion is that if one person held a job position that is on the subtree branching from another person's job position, and the periods during which they held these two job positions overlap, then the first person is said to be a subordinate of the second person. The inference rule is as follows.

$$\text{Job}(\text{Person}_A, \text{Position}_1; T_1) \wedge \text{Job}(\text{Person}_B, \text{Position}_2; T_2) \wedge \text{Successor}(\text{Position}_1, \text{Position}_2) \wedge \text{overlaps}(T_1, T_2) \Rightarrow \text{Subordinate}(\text{Person}_A, \text{Person}_B; \text{intersection}(T_1, T_2))$$

Figure 12 shows the results derived from the deductive inference process.

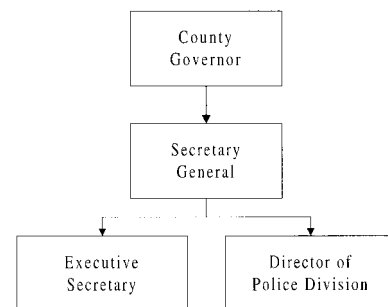


FIG. 11. Bureaucratic hierarchy of Taiwan provincial government during the Chin dynasty.

Key person: Shen, Chao-pin/County governor
 Subordinate: Chao, Shing-ke/Secretary-general
 Subordinate: Chang, Yuan-chu/Executive secretary
 Subordinate: Chie, Weng-da/Director of police division
 Key person: Jiang, Hsiang/County governor
 Subordinate: Chao, Shing-ke/Secretary-general
 Subordinate: Chang, Yuan-chu/Executive secretary
 Subordinate: Chie, Weng-da/Director of police division
 Key person: Wang, Chao-sheng/County governor
 Subordinate: Chao, Shing-ke/Secretary-general
 Subordinate: Chang, Yuan-chu/Executive secretary
 Subordinate: Chie, Weng-da/Director of police division
 Subordinate: Chen, Jia/Secretary-general
 Subordinate: Kao, Lang/Executive secretary
 Subordinate: Chang, Weng-mo/Director of police division
 Subordinate: Chang, Yuan-ying/Secretary-general
 Subordinate: Kao, Lang/Executive secretary
 Subordinate: Chang, Weng-mo/Director of police division
 Key person: Wang, Shi-june/County governor
 Subordinate: Wang, Li-chong/Secretary-general
 Subordinate: Lee, Ting-quai/Executive secretary

FIG. 12. Results derived by the knowledge discovery process.

There are several occasions in which extensive search of the implied relationships among objects in the digital library can be applied to complement human beings' capabilities. One occasion is when a new object is to be added to the digital library. In such an occasion, the search can start with known facts about the new object's relations with the existing objects in the digital library. The implied relations identified by the extensive search provide a clear picture about how the new object is related to the existing objects in the digital library, and are essential for maintaining the integrity of the digital library contents. For example, if the new object is an ancient aboriginal leader, then the search can start with the village where the aboriginal leader resided, and the major events that the aboriginal leader were involved. An extensive search of the implied relations could identify the aboriginal leader's relations with other important historical figures living in the same era.

Another occasion in which an extensive search of implied relations among objects will be invoked is when a content specialist finds a new relationship between two objects and wants to find out whether the new relationship affects our interpretation about existing objects or implies new relationships among existing objects. For example, a newly found piece of information reveals that a person named Y is actually a child of another person named X, and it is already known that a person named Z is a child of X, then a sibling relationship exists between Y and Z.

The extensive search mechanism described in this article can also be applied in informational retrieval. The search mechanism will yield high recall rates, and can rank the search outputs according to the level of association that an output has to the original search target. The power of this search mechanism is that it can find extended relationships among cognitive objects from the existing relationships in the digital library. As mentioned earlier, such extended associations often lead to new interpretations about the subjects of interest.

System Implementation and Evaluation

Figure 13 depicts the system architecture of the NTUDLM. The agent and the objects repository run on a SUN Enterprise 3000 server with four 200-MHz Ultrasparc CPUs, while the content repository is installed on a HP NetServer with a 233-MHz Pentium II CPU. The reason behind incorporating two different platforms is to exploit the advantages of these two different types of platforms. As of today, UNIX-based platforms in general offer higher system performance and higher availability than Windows-NT based platforms. Because the agent and the objects repository are the modules that directly respond to users' queries, it is more desirable to install these two modules on a UNIX-based platform. On the other hand, software utilities for Windows-NT-based platforms are more widely available and at lower costs than those for UNIX-based platforms. As a result, it is more appropriate to install the content repository module on a Windows-NT-based platform. In the implementation, a Microsoft SQL server was employed to build the content repository.

As of today, the NTUDLM contains over 27,000 objects, and the total content volume in the content repository exceeds 400 Gigabytes. The uncompressed image files of the objects account for most of the volume. To expedite transmission of images over networks, the image files are resampled to reduce their resolution, and are compressed before they are loaded into the objects repository. With resampling and compression, the total volume of the image files in the objects repository is about 5 Gigabytes.

With the current volume of the NTUDLM, the compilation process that automatically establishes crossreferences among the Web pages of the objects takes 1,583 seconds. This amount of time does not include the time taken by resampling and compressing image files. Among

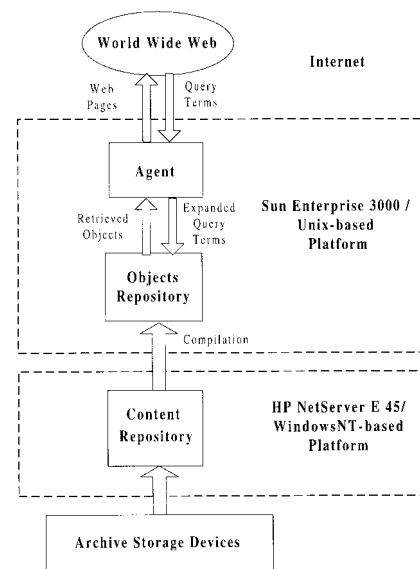


FIG. 13. System implementation of the NTUDLM.

the 1,583 seconds, 1,026 seconds are used to retrieve data from the Microsoft SQL server database, and the remaining 557 seconds are used to build crossreferences among the Web pages of the objects. To evaluate the response time of the system, 10,000 query terms were randomly selected from the content repository primary key list. The system, on average, required 0.157 seconds to service a query.

Conclusion

This article discussed the design of the National Taiwan University Digital Library and Museum (NTUDLM), which addresses content and knowledge management. The digital library design features two major distinctions: (1) the system incorporates a two-tier repository system to facilitate content management, and releases the system manager from manually maintaining hyperlinks among the web pages. (2) The system incorporates an object-oriented model to facilitate the management of temporal information, and exploits information extraction and deductive inference to derive implied knowledge based on the content of the digital library. The incorporated knowledge-processing utilities not only releases specialists from performing simple but tedious inference work, but also extends the applications of digital libraries into a new dimension.

This article also discussed the implementation and evaluation of the NTUDLM. The entire system was implemented on two platforms—one UNIX-based platform, and one Windows-NT-based platform. The main reason behind this practice is to exploit the advantages of two different types of platforms. As far as system performance is concerned, operational experiences revealed that the system can respond to a user's query within a reasonable period of time.

What this article presents are just a few first steps taken by the NTUDLM research team toward developing a knowledge-based oriented digital library. This article focused on deriving the implied relationships among the objects in a digital library containing mostly historical articles and records. Similar ideas can be applied to derive other types of knowledge based on the contents of various types of digital libraries. One issue that is currently under investigation by the NTUDLM research team is the representation and reasoning of spatial knowledge. This study is essential to the goal of incorporating cultural atlas materials into the NTUDLM.

References

- Allen, J.F. (1991). Temporal reasoning and planning. In *Reasoning about plans* (pp. 1–68). San Mateo, CA: Morgan Kaufmann Publishers, Inc.
- Arms, W.Y., Blanchi, C., & Overly, E.A. (1997). An architecture for information in digital libraries. *D-Lib Magazine*. [Online]. Available: <http://www.dlib.org/dlib/february97/cnri/02arms1.html>.
- Baldonado, M., Chang, C.-C., Gravano, L., & Paepcke, A. (1997). Meta-data for digital libraries: Architecture and design rationale. In *Proceedings of ACM digital libraries '97*. Philadelphia, PA.
- Chen, H. (1990). A knowledge-based design for hypertext-based document retrieval system. In *Proceedings of the international conference and workshop on database and expert systems applications (DEXA 1990)*. Vienna, Austria.
- Chen, H., Smith, T.R., Larsgaard, M., Hill, L.L., & Ramsey, M. (1997). A geographic knowledge representation system for multimedia geospatial retrieval and analysis. *International Journal on Digital Libraries*, 1(2), 132–152.
- Chen, H., Yim, T., Fye, D., & Schatz, B.R. (1994). Automatic thesaurus generation for an electronic community system. *Journal of American Society for Information Science (JASIS)*, 46(1), 52–59.
- Cowie, J., & Lehnert, W. (1996). Information extraction. *Communications of the ACM*, 39(1), 80–91.
- Daniel, R., Jr., & Lagoze, C. (1997). Distributed active relationships in the Warwick Framework. In *Proceedings of the second IEEE metadata conference*. Maryland.
- Fagin, R., Halpern, J.Y., Moses, Y., & Vardi, M.Y. (1995). *Reasoning about knowledge*. Cambridge, MA: The MIT Press.
- Galton, A. (1987). *Temporal logics and their applications*. New York: Academic Press.
- Gladney, H.M., Fox, E.A., Ahmed, Z., Ashany, R., Belkin, N.J., Lesk, M., Tong, R., & Zemankova, M. (1994). Digital Library: Gross structure and requirements. In *Proceedings of digital libraries '94*. College Station, TX. [Online]. Available: <http://www.csd.tamu.edu/DL94/paper/fox.html>.
- Glymour, C., Madigan, D., Pregibon, D., & Smyth, P. (1996). Statistical inference and data mining. *Special Issue of the Communications of the ACM on Data Mining and Knowledge Discovery*, 39(11), 35–41.
- Goralwalla, I., Leontiev, Y., Özsu, M., & Szafron, D. (1997a). Modeling temporal primitives: Back to basics. In *Proceedings of 6th international conference on information and knowledge management (CIKM'97)* (pp. 24–31). Las Vegas, NV.
- Goralwalla, I., Özsu, M., & Szafron, D. (1997b). A framework for temporal data models: Exploiting object-oriented technology. In *Proceedings of 1997 conference on technology of object-oriented languages and systems (TOOLS USA 97)*. Santa Barbara, CA.
- Imielinski, T., & Mannila, H. (1996). A database perspective on KDD. *Special Issue of the Communications of the ACM on Data Mining and Knowledge Discovery*, 39(11), 58–64.
- Kahn, R., & Wilensky, R. (1995). A framework for distributed digital object services (Tech. Rep.). CNRI. [Online]. Available: <http://WWW.CNRI.Reston.VA.US/home/cstr/arch/k-w.html>.
- Miller, U. (1997). Thesaurus construction: Problems and their roots. *Information Processing & Management*, 33, 481–493.
- Nuernberg, P.J., Furuta, R., Leggett, J.J., Marshall, C.C., & Shipman, F.M., III. (1995). Digital libraries: Issues and architectures. In *Proceedings of digital libraries '95*. Austin, TX, USA. [Online]. Available: <http://csdl.tamu.edu/DL95/papers/nuernberg/nuernberg.html>.
- Park, Y.C., & Choi, K.S. (1997). Automatic thesaurus construction using Bayesian networks. *Information Processing & Management*, 32, 543–553.
- Russel, S.J., & Norvig, P. (1995). *Artificial intelligence: A modern approach* (pp. 151–184). Englewood Cliffs, NJ: Prentice-Hall International, Inc.
- Schatz, B.R., & Chen, H. (1996). Building large-scale digital libraries. *Computer theme issue on US Digital Library Initiative*.
- Schatz, B.R., Johnson, E.H., Cocharane, P.A., & Chen, H. (1996). Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval. In *Proceedings of ACM digital libraries '96*. Hyatt Regency Bethesda, MD.
- Schutze, H., & Pedersen, J.O. (1997). A cooccurrence-based thesaurus and two applications to information retrieval. *Information Processing & Management*, 33, 307–318.

- Solderland, S., Fisher, D., & Lehnert, W. (1997). Automatic learned vs. hand-crafted text analysis rules (Tech. Rep.). CIIR. [Online]. Available: <http://cobar.cs.umass.edu/pubfiles/te-44.ps>.
- Solderland, S.G. (1996). CRYSTAL: Learning domain-specific text analysis rules (Tech. Rep.). CIIR. [Online]. Available: <http://cobar.cs.umass.edu/pubfiles/te-43.ps>.
- Tong, W.C., Wüthrich, B., & Sankaran, K. (1995). A temporal and probabilistic, deductive and object-oriented query language. In *Workshop on Temporal reasoning in deductive and object-oriented database* (pp. 61–69). Singapore.
- Vila, L. (1994). A survey on temporal reasoning in artificial intelligence. *AI Communications*, 7(1), 4–28.
- Weibel, S. (1995). Metadata: The foundations of resource description. *D-Lib Magazine*. [Online]. Available: <http://www.dlib.org/dlib/July95/07weibel.html>.
- Wong, S.T.C., Hoo, K.S., Knowlton, R.C., Hawkins, R.A., Laxer, K.D., & Tjandra, D. (1997). Issue and applications of digital library technology in biomedical imaging. *International Journal on Digital Libraries*, 1(3), 209–219.