NASA

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
Architectural Design for the Global Legal Information Network

5. Report Date

6. Performing Organization Code

7. Author(s)
Konstantinos (Kostas) Kalpakis

8. Performing Organization Report No.

10. Work Unit No.

9. Performing Organization Name and Address
University of Maryland Baltimore County
302 Administrative Building, 1000 Hilltop Circle
Baltimore, Maryland 21250-5394

11. Contract or Grant No. NAS5-32337
USRA subcontract No. 5555-069-60

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Washington, DC 20546-0001
NASA Goddard Space Flight Center
Greenbelt, MD 20771

13. Type of Report and Period Covered Final
March 1997 - June 1999

14. Sponsoring Agency Code

15. Supplementary Notes
This work was performed under a subcontract issued by
Universities Space Research Association
10227 Wincopin Circle, Suite 212
Columbia, MD 21044 Task 69

16. Abstract

In this report a summary of activities regarding the goals, requirements analysis, design, and prototype implementation for the Global Legal Information Network, a joint effort between the Law Library of Congress and NASA.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
global legal information	Unclassified--Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	1	

NASA Form 1626 Oct 86

Architectural Design for the Global Legal Information Network

Konstantinos Kalpakis
Computer Science and Electrical Engineerign Department
University of Maryland Baltimore County
1000 Hilltop Circle Baltimore, MD 21250
email: kalpakis@csee.umbc.edu

Summer 1998; Updated on Summer 2000

Abstract

In this report, we provide a summary of our activities regarding the goals, requirements analysis, design, and prototype implementation for the Global Legal Information Network, a joint effort between the Law Library of Congress and NASA.

Contents

1 Global Legal Information Network: Overview **5**

1.1 Vision 5

1.2 Scope and Capabilities of GLIN 7

1.3 Development Approach 8

1.4 Technology Issues 10

1.4.1 Standards 10

1.4.2 Database Technology 11

1.4.3 Legal Metadata 11

1.4.4 Indexing, Query, and Searching 11

1.4.5 Collaboration Services 12

1.4.6 Currency and Temporal Evolution 12

1.4.7 Billing and Charges 12

1.4.8 Security and Authenticity 13

1.4.9 Accessibility and Connectivity 13

2 GLIN System Architecture **14**

2.1 Functional and Performance Requirements 15

2.2 Architecture 17

3 Prototype **20**

3.1 Database and Relational Schema 20

3.1.1 Document Type Definition for GLIN 23

3.1.2 Relational Schema for Heterogeneous Legal Documents 30

3.1.3 XSL for GLIN Document Presentation 31

3.1.4 Indexing 38

3.1.5 Query Services 43

3.1.6 Security and Authentication 43
3.1.7 Online GLIN Document Maintainance 43

Appendix A: Summary of Activities for 1996–1997 **44**

Appendix B: Summary of Activities for 1998 **47**

List of Figures

1 Servicing requests in an agent–based architecture. 19

2 Subsystems in the GLIN System Architecture. 48

3 GLIN Prototype System. 49

4 User specifies query to search for GLIN documents related to "rivers". 50

5 Results of user query. 51

6 User requests to see metadata for the 1st hit. 52

7 User requests to see the generic gist of the 1st hit. 53

8 User requests a thematic list to be generated and then requests the gist of the 1st hit from the "harborage" point of view. 54

9 User requests to see the 1st hit with the terms maching his/her query highlighted and hyperlinked with each other. 55

10 User requests to see the full-text of the 1st hit. 56

11 User uploads a local file to the remote GLIN server. 57

12 User edits a record for a GLIN summary document. 58

13 User edits a record for a GLIN summary document. 59

14 User edits a record of the publications table. 60

1 Global Legal Information Network: Overview

Each country has a body of legal information, including statutes, and regulations which codifies its internal and external operating procedures. There is a growing need in each country to rationalize, and among countries to harmonize these bodies of legal information. A requirement, therefore, may be deemed to exist for an international networked computer information system, which can provide the necessary global legal information services. International cooperation can be greatly enhanced by such a system which can make relevant local laws and regulations mutually accessible to countries engaged in, or considering, such joint ventures. Examples of the laws of interest are those governing trade, narcotics, and fire arms. A powerful distributed information system is very crucial to fulfilling such a requirement.

The Law Library of Congress has developed a basic viable prototype, known as the Global Legal Information Network (GLIN), for the acquisition, processing and retrieving of digitized texts. However, the demand for access to information on the statutes and regulations of nations concerning a wide range of subjects is of considerable volume and complexity. In order to facilitate maintaining and searching such large databases with adequate flexibility and speed, advanced digital library technologies must be sought and integrated into GLIN. Further, communications requirements have to be investigated and supported.

Working with academia and industry, researchers at NASA have been closely following and advancing scientific data management technology. Further, they have been applying such advanced research findings to efficiently manage large volumes of satellite data, in support of earth and space scientific investigations. Those same research concepts and technologies can be transferred to enable GLIN to serve and expand its potential user community, while meeting its requirements.

This effort is aimed at exploiting the unique synergism of our team which embodies researchers from the Law Library of Congress, NASA, academia and industry in order to provide a distributed, easy to use, expandable, flexible, and efficient implementation of GLIN based on the state-of the-art in information technology. This project has two specific goals designed to ensure that GLIN can continue to sustain the rapid increase in demands on this service. Starting from the current state of GLIN, this project is intended to: 1) upgrade the GLIN system to take advantage of the latest developments in technology, and 2) enhance the technological infrastructure of GLIN in order to allow for future enhancements and expansions of GLIN's functionality and geographical coverage.

1.1 Vision

The demand for access to basic information on the statutes and regulations of nations concerning a wide range of subjects is of considerable volume and complexity. This demand

may be stated to be of a global character. Such access is of little value if accuracy and currency may not be assured. Accuracy and currency in a global context may presently be obtained by digitization and networking. Accuracy is secured by accessing the official standard sources of publication. Currency requires that amendments and repeals as well as new enactments or issuances be accessed in a timely routine basis.

The Law Library of Congress has considerable experience in effectively managing this kind of information and developed a basic viable testbed for the acquisition, processing and retrieving of digitized texts. However, additional technological support to achieve acceptable capabilities is still needed.

It is the project's goal and vision that:

- GLIN provides an appropriate means of security to guarantee the authenticity of the producer, requester, sender and receiver of legal data, as well as that of the data itself.
- GLIN provides sufficiently high bandwidth capabilities to insure interactive usage of the database as a digital legal library anywhere in the world.
- GLIN operates on a technology infrastructure that supports current international or de facto standards to provide basic access to any country of the world desirous of participation.
- GLIN trains member countries and develops international regional training centers.
- GLIN is established as an international entity for the purpose of establishing GLIN policy and managing its resources.
- GLIN has established the means and the relationships to be financially self-sufficient in sustaining the capabilities to receive and transmit digital data and information by whatever means are appropriate, including dedicated and leased bandwidth.
- GLIN has acquired the skills and experience to recommend organizational, procedural and funding models for participating countries.

GLIN offers direct benefits to the members of the network in at least the following:

- Participation in a state-of-the-art global electronic network.
- Participation in research and development through demos of advanced processing and networking technologies.
- Developing a national electronic legal information system in the original official language for local users, and in English for global users.

- Access to the statutes and regulations of own and all other countries in the network.
- Training and technical assistance as required.
- Shared design, development and maintenance of the network at minimum cost per member.

1.2 Scope and Capabilities of GLIN

GLIN as a legal information system, can help each country assess the effectiveness of its own laws and provide a research tool to aide future legislations. GLIN can also enhance and expedite business transactions and international collaboration in numerous areas among member countries. Furthermore, with the expected growth in membership, GLIN can help international and regional organizations resolve international disputes or form commercial alliances. These are just a few examples of the potential benefits of GLIN.

In view of the above, the Law Library of Congress made an assessment of capabilities available to expand its existing International Legal Database, consisting of abstracts and index terms related to statutes and regulations extracted from official sources from some thirty countries in the Americas, Europe and Africa, into a global system intended to include a much larger number of countries as well as access to the full texts of the instruments concerned. This task requires participation of resources beyond those available to the Law Library of Congress itself.

Based on the expressed high interest of governments, the global business community, and others on current, authentic legal information, the concept of inviting national law making bodies to participate on a cooperative electronic network emerged. The first two nations to express their interest were Brazil and Mexico. They were invited to participate with the Law Library of Congress (GLINCENTRAL) in the initial efforts of testing the basic concepts and elements of the GLIN project. These three stations constituted the cornerstone of the organization that soon was christened the Global Legal Information Network. Based on the results and experiences of their work the Law Library of Congress launched the testbed. The GLINSTATION was cast including hardware, software and personnel as well as a body of written specifications and procedural standards to be observed in:

- identifying authentic sources;
- selecting instruments;
- analyzing and abstracting the instruments;
- building thesaurus for validation of descriptors;
- capturing and digitizing texts;
- data inputting;

- transmission and reception of data;
- organizing and indexing the data and
- search and retrieval.

A set of principles designed to govern the rights and duties of the members has been drafted. Its approval by members is in progress.

The first target of GLINCENTRAL has been to establish GLINSTATIONS in the participating countries. In addition to the three original members, seven countries have been accepted for membership. Training of their designated staff has been completed at GLINCENTRAL. A number of other countries have made formal statements of interest and their admission is being negotiated. The current list of countries that are either a part of GLIN or have indicated interest includes Argentina, Paraguay, Uruguay (in the Americas); Hungary, Poland, Lithuania, Ukraine (Europe); Israel, Kuwait, S. Korea (Asia); Mauritania, Morocco, and Egypt (Africa).

Most countries currently admitted are in the process of acquisition of the recommended standard hardware, software and telecommunication capabilities. However, their trained staff are actively engaged in complying with organizational and basic information processing routines. GLINCENTRAL has in operational mode the following capabilities:

- An established model consisting of a set of standards and procedures for the analysis of legal instruments, formulation of the corresponding abstracts and thesaurus building.
- Training curricula and materials for the legal and technical personnel designated by the member countries.
- Transmission and reception of digital data via the GLINCENTRAL Internet node.
- Storage of and controlled access to the data at the GLINCENTRAL server.

In addition, and architecture has been selected and used for the GLIN client station, with the minimum requirements described elsewhere.

1.3 Development Approach

Legal data when stored in digital form may become a passive collection of data which would be cumbersome and slow to work with. GLIN's goal, however, is that data should be accessible in interactive and flexible ways so that the amount of work done by lawyers and clerks is reduced. Therefore, GLIN can be viewed as a digital legal library carrying out and

expanding its functions by incorporating advanced methods of storage, classification, and searching in order to provide a set of smart services over the stored legal data.

The GLIN development approach proposed here follows a two-tiered plan: (1) Upgrade, and (2) Enhance. In this plan, we start on the short run (upgrade) by augmenting GLIN with compatible state-of-the-art technologies in acquisition, communications, and databases in order to upgrade the current system.

A longer-term plan (Enhance), however, will include establishing strong infrastructure based on advanced research concepts, which can sustain the potential future increase in functionality and geographical coverage. The specific enabling concepts and research issues include:

- Legal data will be more useful to lawyers and clerks if it can be retrieved based on semantic inquiries. Otherwise the user has to manually search and browse the database which is time consuming. In order to search the data with semantic concepts we need to index the data first. This requires content-based indexing methods. The data is indexed according to the content, possible use and context of the data which enables queries to be processed against such indexes.

- In a distributed data environment there is always the problem that different data use different technologies and platforms. If we want easy declarative access to distributed information we need to either enforce standards that are followed by all distributed stations or we need inter-operating software that hides these implementation details from the user.

- We need to have the appropriate database technology to store the large amount of data that is expected in GLIN. Moreover, the retrieval engines need to be optimized so that the performance of GLIN is acceptable when large amounts of data are stored and retrieved. We also need to ensure that the database technology that is used has adequate recovery and logging facilities so that usage can be logged and analyzed and recovery from crash failures can be guaranteed.

- In a distributed GLIN environment when the bandwidth becomes a problem we need to build replicated or mirror copies of frequently accessed data electronically closer to the users. This reduces network resource utilization resulting in improved network throughput performance. The database technology used in GLIN and the distributed data protocols need to work with the replicated and mirror copies of data.

- Careful attention has to be paid to provide global connectivity in GLIN to all members. We anticipate high-bandwidth connections to some member sites. These connections have to be highly-available possibly on 24x7 basis. In cases of network failure, members will be provided alternative access sites possibly conatining replicated data repositories. For some members in outlying areas satellite communications may be used to provide the necessary connectivity and bandwidth.

- Access to GLIN data needs to be monitored and authenticated. Members need to be assured of the integhrity of the data. This means that changes and additions to the data can only be done by authorized users in formally accepted procedures.

Once a globally accessible GLIN database has been made available we can define a host of advanced services on top of this flexible database to reduce the workload of lawyers and clerks. Some possible such services are described below.

- Locator–service finds related precedents; the finder box allows the user to type a set of circumstances and contexts through a set of sentences and the finder finds all relevant precedents and rank orders them.
- Navigator–service allows interactive navigation by following citation links in legal text, allows a visual map to be constructed of the citations that have been navigated. This map can be saved and mailed to other users, thus sharing what one user has found with other users.
- Notetaker–service allows users to add margin notes to text that has been read and accessed by other users. In this way users will be able to find out what other users have been accessing, what their thoughts are about a particular subject. This note taking facility can be viewed as an abbreviated form of forum discussions and electronic meeting rooms. The interesting part of this proposal that makes it very useful is that the notes and thoughts of users are tied to the legal text which makes the notes contextual and relevant. Only users who are reading a certain text will see these notes; all other users will not be subjected to irrelevant information.
- Collaborator–service allows visual maps of navigation and notes to be made available to colleagues for collaborations. A user will be able to take all notes and navigation links about a certain topic and mail them to a collaborator. In this way. members and can collaborate on areas of mutual interest.
- Discusser–service allows queries to be posed on a text that has been accessed; the queries are sent of to virtual discussion rooms where other people who have accessed the same document visit, and who can respond to the queries.
- Analyzer–service allows statistical analysis of legal data contained in GLIN

1.4 Technology Issues

1.4.1 Standards

Among the most significant information management issues is the task of establishing standards. These standards can be at either of two different levels: standards for data interchange

or standards for interoperability. The latter standard is the more problematic to establish. In practice, however, where there is either more than one standard when simply one would suffice, or where there is no standard at all where one is necessary, de facto industry-centered standards should be taken into consideration. In the area of document description in GLIN, for example, SGML (Standard Generalized Markup Language) is an attractive standard because it has a dual-purpose language suitable for both paper and electronic publishing. Furthermore, search engines could take advantage of its markups to enhance retrieval performance. Whatever the case may be, one will need to support plain ASCII, Postscript, and TIFF, a popular format for digitized text.

1.4.2 Database Technology

Another information management issue is database technology, which is crucial on the data storage and management level. Extended relational DBMS (SQL 3-based) or object-oriented DBMS can be adopted and specialized technologies for managing temporal, spatial, and geographic information should be identified and incorporated.

1.4.3 Legal Metadata

There is also a need to provide mechanisms for uniformly accessing heterogeneous information sources which cannot all be assumed to be homogeneous. Thus, it will be necessary to provide meta-information or meta-data, such as a source model, about these information sources. Such meta-data should be based on the content of the information available at a source, on the creation and maintenance of the information at the source, on the format and schemas of the information maintained by the source in accessing its information resources. It is necessary, however, for that purpose to determine suitability of the IDL (Interface Description Language.)

1.4.4 Indexing, Query, and Searching

Other information management issues include text analysis and information retrieval methods for converting, indexing, representing, searching, and presenting the desired information. Yet there must also be some focus on human-computer interaction paradigms to help users effectively learn, search, and utilize the information available in digital law libraries. There will also be a need to provide a mechanism for query refinement because the user is usually unable to effectively pose a query that reflects his or her information needs through the use of a traditional information retrieval system. The process can be construed as an iteration having the following phases:

submitting a query, obtaining search results, analyzing their relationship to the query, reformulating the query, starting over again.

Queries should be based on such search technique as keyword search, attribute-based search, similarity to previous queries, proximity and content-based search. This can be facilitated by adopting powerful legal metadata and the use of database technology that can take good advantage of it. Indexing of documents is ultimately a crucial information management issue. Techniques for automatically resolving terminological differences will be beneficial to the users of GLIN. Thus, there will be a need to develop domain-specific technologies, and methods for resolving domain and scope differences, as well as synonyms and antonyms. Natural Language Processing and Knowledge-Based Techniques will be helpful in the tasks of abstracting and summarizing. Latent-semantic indexing techniques may also be useful to develop a teachable indexing mechanism that takes similar concepts into consideration when doing a retrieval.

1.4.5 Collaboration Services

GLIN should provide such mechanisms as electronic publishing tools to support communication and collaboration among its users and to support sharing of annotations and sets of documents among specialized groups of users or expert groups.

1.4.6 Currency and Temporal Evolution

In the event that the documents in a digital law library change rather frequently in their lifetime, we foresee the need for a version management system. Such a system should enable a user to examine the evolution of a given document and it should possess the capability to be used for managing amendments to existing documents. Temporal database technology could be used to accommodate such needs.

1.4.7 Billing and Charges

Since a GLIN user will invoke may services in accessing information resources cost is a matter which must also be considered. Services should be priced individually and at low cost. Consequently, there will be a need for an Electronic Payment System, electronic cash, or electronic credit cards. Moreover, to sustain growth and to promote private investment in GLIN, a fee collection service must be established.

1.4.8 Security and Authenticity

Due to its sensitivity to interpretation, legal information in GLIN must be authentic. Maintaining authenticity is a challenging requirement, however, as it relates to the fidelity of acquisition, data format, as well as system security. Technological guidelines need to be established to offer secured access to authentic information in GLIN, with good readability and viewing flexibility.

1.4.9 Accessibility and Connectivity

Variety in the number of ways to access GLIN is another information management concern. A user should be able to access GLIN through a variety of terminals, starting from plain VT 100-type of terminals to sophisticated terminals with advanced graphics capabilities. To further maximize usage of GLIN, it would be prudent to support access through a touch-tone phone and a teletype or fax, through an old generation PC and a slow modem, through an advanced workstation with a high speed connection to the Internet. When accessing GLIN special attention is required to make sure that performance bottlenecks are avoided. This problem will only be intensified by the high bandwidth requirements for transmitting documents as images (TIFF format) rather than as encoded text. It is of utmost importance to avoid bottlenecks in order for GLIN to achieve significant growth and be of maximum value to its users.

Furthermore, to improve performance and reliability, GLIN should support replication of information. For example, one can start by using existing software for mirroring archive FTP sites Straying from the information management issues, connectivity for GLIN creates another set of issues which need to be considered. Namely, there is a concern that there are still sites that are interested in participating in the GLIN experience, but which are not connected to the Internet. This is particularly important for developing countries in Latin and South America, Africa and Asia. Consequently, increasing connectivity to such countries is an important goal.

One approach for increasing connectivity is to use technology developed for Fidonet. Fidonet is a point-to-point and store-and-forward wide area network utilizing modems and dial-up phone lines. It provides low-cost connectivity among individuals, while trying to minimize modem time. As of April 1993, there were about 20,000 Fidonet nodes around the world, over 75% of them in North America and Europe, with less than 10% in Asia, Africa, and Latin America. Fidonet is popular among amateur system operators like BBS operators. Software implementations are available to port it on a variety of PCs and other systems. The primary function of Fidonet is forwarding news and exchanging email messages. There are also gateways connecting Fidonet networks with the Internet.

SLIP and PPP provide another alternative for increasing connectivity. SLIP (Serial

Line Internet Protocol) is a protocol for Internet point-to-point connectivity via modems over telephone lines between two systems. PPP (Point-to-Point Protocol) is another direct link protocol which works over serial lines and direct links similar to SLIP. Implementations for both are available over a variety of PC and Workstation platforms.

Whatever the case may be, GLIN should support a variety of transport protocols like X.25, TCP/IP and Windows Socket API Winsock, BITNET, UUCP, HTTP, Mobile IP, SLIP, and PPP. GLIN should also support protocols for packet radio communications (low bandwidth), exchange of information via telephone only (a touch-tone phone and speech/text translators), teletype and fax. It is also necessary to experiment using a combination of CD-ROM technology together with standard internetworking to see whether these methods maximize efficiency and more effectiveness, both in terms of access performance and in terms of costs. Documents that seldom change and have reasonably extensive lifetime, for example could be distributed via CD-ROMs to regional GLIN servers where users can access them while amendments, repeals, and documents that change frequently are distributed over the Internet. Satellite communications can be used to augment the bandwidth available and support countries that have no internet access. At GSFC extensive experience with using the ACTS satellite for distributed information processing has been developed.

Still, there are more practical information management issues which are inevitable in a comprehensive discussion of the GLIN project. Among such issues is the matter of copyright and intellectual property rights both of which can be dealt with appropriately by the Library of Congress' through the U.S. Copyright Office and the Library of Congress' involvement in the ARPA sponsored Electronic Copyright Management System. Any potential obstacle regarding authenticity of information, both for provided documents and annotations to them by users, can be resolved through the use of electronic signatures based on public-key cryptography. For user authentication, cryptographic techniques such as smart-cards and electronic pockets could be an option. Not yet resolved are issues regarding export controls and usage of cryptography by various participating countries.

2 GLIN System Architecture

The high–level objectives set forth for the GLIN System are:

- The system should be attractive to users and to content and value–added–service providers.
- The system should be an integrated solution to the problem of managing and manipulating legal multi–lingual instruments for the Library of Congress and member countries.
- The system should be operational for a long–period of time.

- The costs of building, operating, and maintaining the system should be kept under control, and in no case should outperform the benefits.

2.1 Functional and Performance Requirements

The GLIN system should provide for the following services:

- Efficient, Flexible Indexing and Retrieval. Supporting storage and retrieval of a large collection of digitized indexed objects presents challenges that are not so critical in small ($\approx 1 - 3$ GB) text–based databases. Flexibility deals with the ability to handle a large variety of formats and codings for the data objects.

- Query Formulation Assistance and Term expansion. This is critical in order to address differences on the user's and system's vocabularies and make the system more attractive to the user. Reliability, efficiency, and self–diagnosis are important criteria of success.

- User Interfaces. Should support query formulation, presentation of retrieved data, feedback, browsing, collaboration, and sharing. The challenge is to give the user sophisticated functionality in an intuitive and simple manner.

- Monitoring, Routing, and Filtering. The idea here is to allow the user to monitor an information stream and present to that user those data items that he/she had specified in a complex user profile. This is very important in government and industrial settings.

- Effective Retrieval. Deals with the evaluation of relevance of the retrieved data to the user's query. Standard measures of effectiveness are precision and recall. The system should be able to assess the degree of errors and mismatches between expected results and retrieved data (e.g. self-assessment of false positive and false negative error rates). To this end, models that capture the user's perceived relevance or similarity between data objects should be investigated.

- Workflow management for depositing and maintaining data. Provide an integrated flexible workflow management environment to handle and/or automate the processes of depositing new data and maintaining stored data.

- Distribution, Replication, and Consistency. Provide for ways to balance load of user's requests based on the requested quality of service, and provide for the distribution and replication of data items in a wide–area network of servers while maintain the consistency and integrity of the data.

- Multimedia Capture, Storage, Indexing, Retrieval, and Delivery. Handling language related multimedia, such as text in images, images in text, digitized scanned documents, speech, and video. Special methods and techniques need to be developed to

deal with the whole range of issues from capturing a non–text data item, to delivering it to the user.

- Information Extraction. Capability to extract from the stored data such things as entities in the data, their attributes and other features in the data, together with their relationships. For example, extracting references to legal instruments, institutional entities mentioned in data. The data items could be presented in a variety of formats and codings such as text, document images, audio, video. Techniques from text understanding, pattern recognition, image analysis, and optical character recognition should be investigated.

- Abstracting and Result Integration. Automatically create informative abstracts of stored or retrieved data. Support the ability to integrate data retrieved from multiple sources, ranked by each source individually, in an integrated ranked way to the user.

- Feedback and Iterative query refinement. Obtain feedback from the user and provide workspaces where the user can refine his/her queries in an iterative way.

- Information Analysis and Data Mining. Provide support to the users to perform analysis of the information present and study their relationships.

- Annotation and Collaborative Legal Research. Provide the user the ability to annotate data items and share annotations with other users. Provide an environment where legal experts can collaborate remotely in both an asynchronous and a synchronous manner.

- Pricing and Charging. This is a must in order to provide the necessary incentives to create and maintain high quality data items.

- Authenticity, Privacy, and Security. Support methods to authenticate the integrity of the retrieved data. Protect the privacy of the users of the information, while at the same time maintain the integrity and security of the system and provide for charging the users of the system. This is further complicated, by supporting the ability to analyze the usage of the system and its data.

The performance requirements of the GLIN system are as follows:

- Acceptable response times to user's requests with graceful degradation in case of failures or overloading.

- High degree of fault–tolerance and self–stabilization.

- Handle data repositories requiring storage in the 100's GB range.

- Sustain a large peak rate of user requests.

2.2 Architecture

GLIN should have an open system architecture with the following properties:

- Modular. Components can be added, replaced, or removed without affecting the functionality of the other components.
- Extensible. New elements can be easily included and integrated with the existing ones. These could be from new stations, to new services, to new data types and data representations.
- Scalable. The system should be able to handle millions of data objects and users with a tolerable performance degradation. This in turn implies that the architecture will be distributed and support replication.
- Support Heterogeneous and Autonomous Components in a Federation. We accept that heterogeneity and autonomy are both necessary and desirable. Necessary since its almost impossible to impose strict requirements of conformance to every aspect of the system, and desirable since it reinforces the other properties of the architecture in modularity, extensibility, scalability, and fault–tolerance.
- Fault–Tolerant. Graceful degradation in performance with hardware and software failures.

The basic structuring elements of the architecture of the system for GLIN:

- Data Objects (DO). A data object is item that is manipulated in the system. It consists of three parts. The data, the key–metadata, and the metadata. The key–metadata of a DO contains at least the handle for the DO. The data part of the DO is typed. A DO is one of the following types: a bit stream, a character stream, a set or sequence of bit streams, of character streams, of handles of other DOs, of other DOs. Additional types for DOs can be introduced for convenience. The metadata part contains any additional metadata for the DO such as those defined in the Dublin Core Element Description (subject, title, author, etc.) A distinction is made between stored DOs, registered DOs, and derived DOs.
- Data Object Handle (DOH). The DO handle is a globally unique string that identifies a data object. A DO handle could modeled after the Internet draft for Uniform Resource Names. DO handles, like URNs, have the following characteristics: global scope, global uniqueness, persistence, extensibility, and independence. A DO has exactly one DO. A DO may be read–only or read–write.
- Data Object Handle Server (DOHS). Data Object Handle generation, resolution, and mapping is done by a a DOH server. The originator of a DO obtains a handle for that

DO from a DO Handle generator. The handle generator is responsible for guaranteeing uniqueness of the handle. DOH servers are organized in a hierarchy. In order to manipulate a DO via its handle, a service that maps DO handles to their locations is needed. DO locations are modeled after the Internet Draft on Uniform Resource Locators (URL). The resolution and mapping of a DO handle to the location of a DO is achieved through a Handle Server. The Handle servers are modeled after the Internet Draft on Uniform Resource Characteristics (URC) service.

- Data Object Repository (DOP). The data object repository is responsible for storing, providing access, and maintaining data objects. When a data object is stored in a repository, it becomes a stored data object. When a stored data object also has a handle, it becomes a registered data object. Processes access data objects in a repository through a simple protocol that provides the basic data manipulation functions independent of content, e.g. get DO, get DO's key–metadata, get DO's metadata. A user can either access stored data objects directly through that protocol, or can access registered data objects through mediators and other value–added services.

- Data Object Transaction Log (DOTL). Every operation on a registered data object is recorded in a transaction log. This log is available only to authorized processes or users. The storage and maintenance of DOTLs is

- Mediator. A Mediator is a program that collects information from one or more data sources, processes and combines it, and exports the resulting information.

- Facilitator. A facilitator is a program that is performing a complex task by selecting a plan and using a dynamic configuration of other programs to execute that plan.

- Transducer. A transducer is a program that mediates between a request for a certain task and another program executing a similar task that recognizes a different protocol.

The proposed GLIN system architecture is strongly–influenced by agent–based architectures. Figure 1 shows a general case of servicing a client's request in an agent–based architecture. The client contacts a facilitator(s), which selects the plan to use in servicing the request, recruits the relevant other "agents", and executes the plan. In accessing a repository, certain transformations need to be done either to the request or to the data retrieved.

Figure 2 shows the subsystems of the GLIN system architecture, together with the central interactions among them. The functionality of the subsystems shown in figure 2 is as follows:

Input Subsystem: Capture and deposit data into GLIN. Data objects include text, digitized text, images, audio, and video. Also, register the data with the handle management subsystem.

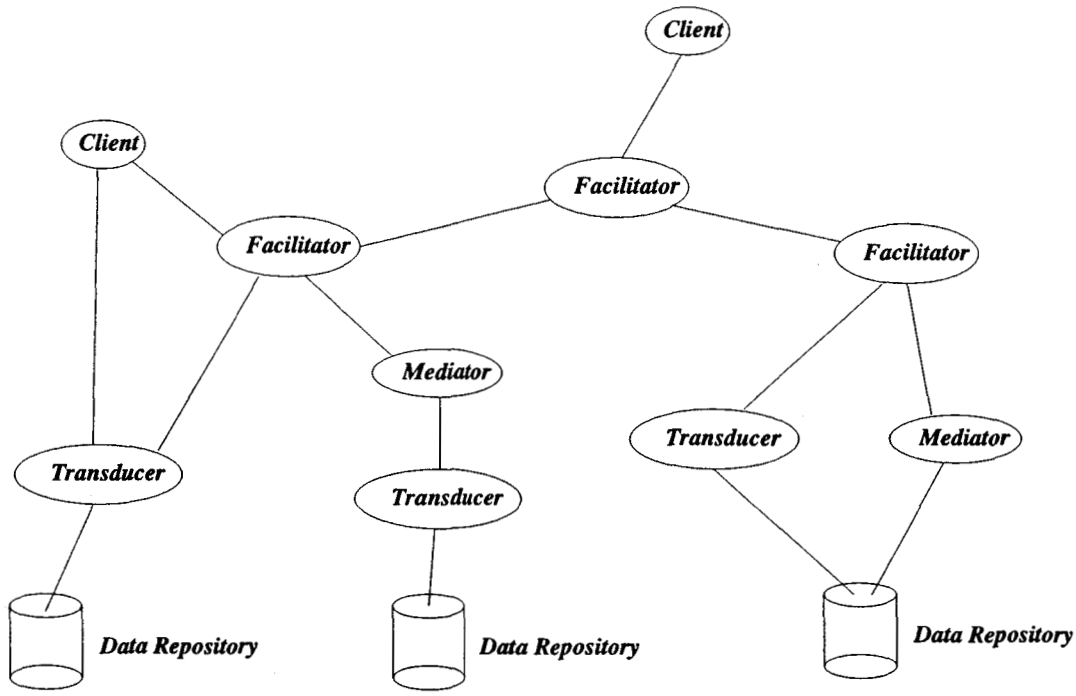

Figure 1: Servicing requests in an agent–based architecture.

Handle Management Subsystem: Generate unique handles for data objects for the GLIN system, map handles to the locations of the corresponding data objects.

Storage Management Subsystem: Provide the necessary storage management functionality for the GLIN distributed data repositories. Also maintain the transaction log for data objects.

Integrity Control Subsystem: Provide functionality for the integrity control of registered data objects and for their authenticity.

Information Extraction and Abstracting Subsystem: Perform the feature and information extraction functions needed for indexing the registered multimedia data.

Indexing Subsystem: Create and maintain the indexing structures for efficient and effective retrieval of the registered data.

Distributed Retrieval Subsystem: Identify the information sources to search, evaluate user's queries to identify relevant registered data. Generate ranking of relevant data in the answer, summarize the answer, and provide self–assessment of quality of the answer.

Data Presentation and Delivery Subsystem: Provide functionality for visualizing and delivering retrieved data to the user.

Pricing and Charging Subsystem: Provide for the pricing of the services provided and information requested, and for charging the user with the corresponding cost. Support for protecting user's privacy.

Monitoring, Routing, and Filtering Subsystem: Provide functionality for evaluating user's profiles against new registered data objects.

Collaborative Workspaces Subsystem: Provide functionality for annotating and sharing annotations among users. Also, provide for collaborative legal research among users.

Query Formulation Assistance Subsystem: Provide assistance to the user in formulating queries against the registered data. Handle differences in vocabulary, do term expansion, query–by–example, process user's feedback, and support iterative query refinement.

Data Analysis Subsystem: Provide for performing various analyses of the registered data and their transaction log records.

User Interfaces: Interact with the user. All interfaces for interacting with the user are provided here.

There are also at least two other subsystems: a workflow management subsystem, providing services for managing the flow of information in GLIN, and the recovery subsystem that handles fault–recovery issues.

It is envisioned that many of the subsystems above will follow the agent–based skeleton shown in figure 1.

3 Prototype

We have created a prototype demonstrating many of the capacibilities and functionalities for GLIN mentioned earlier. This prototype is the basis for another prototype system, ELIS (Environmental and Legal Information Systems) that is under development.

Our GLIN prototype is built using the Apache 1.3 WebServer, Java Server Pages Technology, Java Beans, the Oracle 8.1.5 DBMS, and the Tomcat XMl/XSL processors. A block diagram of the GLIN prototype is shown in Figure 3.

3.1 Database and Relational Schema

The GLIN prototype system uses the Oracle 8.1.5 Object–Relational Database Management System. The schema for the GLIN database is as follows:

```
create table abstract
  (nu_abstract_id number(7) primary key,
   nu_created_by  number(7) not null,
   dt_created     date not null,
   nu_updated_by  number(7),
   dt_updated     date,
   nu_issue_id    number(7) not null,
   nu_class_id    number(7),
   tx_instrument_number varchar2(30),
   nu_pdf_id      number(7),
   dt_issueance   date,
   tx_title       varchar2(2000),
   tx_summary     long,
   nu_provisions  number(5),
   tx_title_searchkey varchar2(50));

create table abstract_terms
  (nu_abstract_id    number(7) not null,
   nu_thesaurus_id   number(7) not null
   nu_seq            number(7) );

create table citation
  (nu_citation_id      number(7),
   nu_citation_type_id number(7),
   tx_citation         varchar(2000),
   nu_language_id      number(7),
   nu_abstract_id      number(7));

create table citation_type
  (nu_citation_type_id number(7),
   tx_citation_type    varchar(50));


create table class
  (nu_class_id number(7) primary key,
   tx_class   varchar2(200) not null,
   nu_country_id    number(7));

create table country
  (nu_country_id    number(7) primary key,
   na_country       varchar2(60) not null,
```

```
   cd_trained        number(1),
   cd_pending        number(1),
   member            number(1),
   na_contributing_facility varchar(200),
   tx_directory      varchar(30),
   cd_internet       char(2));

create table hitlist
  (nu_state_id     number(7) not null,
   nu_seq          number(7) not null,
   nu_abstract_id number(7) not null);

create table issue
  (nu_issue_id number(7) primary key,
   nu_publication_id number(7) not null,
   tx_issue_number   varchar2(200),
   dt_issue    date,
   tx_specifics  varchar2(400) );

create table languages
  (nu_language_id    number(7),
   tx_language       varchar2(200));

create table person
  (nu_person_id number(7) primary key,
   na_first   varchar2(40),
   na_last    varchar2(40)  not null);

create table publication
  (nu_pub_id number(7) primary key,
   tx_pub_title       varchar2(2000),
   nu_country_id  number(7));


create table relationships
  (nu_rel_id number(7),
   tx_active          varchar(30),
   tx_passive         varchar(30) );

create table relationship_status
  (nu_act_abs_id      number(7),
   nu_pass_abs_id     number(7),
```

```
   nu_relationship_id number(7) );

create table role
  (tx_id             varchar2(20)  not null,
   nu_person_id   number(7) not null,
   nu_roletype_id number(7) not null,
   tx_password       varchar2(20)   not null,
   nu_country_id number(7));


create table roletype
  (nu_roletype_id number(7) primary key,
   tx_type   varchar2(40) not null,
   cd_security    char(3)  not null);

create table thesaurus
  (nu_thesaurus_id   number(7) not null,
   tx_term           varchar2(200) not null,
   nu_broader_term   number(7),
   nu_use_term       number(7),
   nu_see_also_term  number(7));

create table pdf_handle
  (nu_pdf_id number(7) not null,
   tx_handle varchar2(512) );
```

Upon loading this database with the collection of GLIN data obtained from the Library of Congress, we discovered a number of inconcistencies in the data. In order to continue with our effort, we undertook the effort to clean up the data and bring the GLIN database into a consistent state.

3.1.1 Document Type Definition for GLIN

To facilitate integration with other sources of legal documents, separate the data coding from the data presentation, and enable effortless customization and client processing, we define an XML Document Type Definition (DTD) for the GLIN data which is as follows:

```
<?xml version="1.0"  encoding="US-ASCII" ?>
```

```
<!-- DTD for GLIN documents
-->

<!ENTITY % languages "CDATA #IMPLIED" >
<!ENTITY % countryCodes "CDATA #IMPLIED" >

<!ENTITY % thesauri "CDATA #IMPLIED" >
<!ENTITY % subjectTermType "(controlled|uncontrolled) 'uncontrolled'" >

<!ENTITY % citationTypes "(primary|alternative) 'primary' " >

<!ENTITY % relationshipTypes "CDATA #IMPLIED" >

<!ENTITY % relationshipRoles "(active|passive) 'active' " >

<!ELEMENT glindoc (docid?,
                   country?, class?, number?,
                   title?,
                   subjectIndex?,
                   issueance?,
                   issue?,
                   summary?,
                   citation*,
                   crossReference*,
                   linkage*,
                   relationship*,
                   docimage*,
                   note*) >

<!ELEMENT docid (#PCDATA ) >

<!ELEMENT country (#PCDATA) >
<!ATTLIST country code %countryCodes; >

<!ELEMENT class (#PCDATA) >

<!ELEMENT number (#PCDATA) >

<!ELEMENT title (#PCDATA) >
<!ATTLIST title language %languages; >

<!ELEMENT subjectIndex (term*) >
```

```
<!ATTLIST subjectIndex type      %subjectTermType; >
<!ATTLIST subjectIndex thesaurus %thesauri; >
<!ATTLIST subjectIndex language  %languages; >

<!ELEMENT term (#PCDATA) >
<!ATTLIST term id    CDATA #IMPLIED >
<!ATTLIST term order CDATA #IMPLIED >

<!ELEMENT issuance (date?) >

<!ELEMENT date (#PCDATA) >
<!ATTLIST date year  CDATA #IMPLIED >
<!ATTLIST date month CDATA #IMPLIED >
<!ATTLIST date day   CDATA #IMPLIED >

<!ELEMENT issue (#PCDATA | publication |number | date | specifics)* >

<!ELEMENT publication (#PCDATA) >
<!ATTLIST publication language  %languages; >

<!ELEMENT specifics (#PCDATA) >

<!ELEMENT summary (#PCDATA) >
<!ATTLIST summary language %languages; >

<!ELEMENT citation (#PCDATA) >
<!ATTLIST citation type %citationTypes; >
<!ATTLIST citation language %languages; >

<!ELEMENT crossReference (#PCDATA | title | linkage)* >
<!ATTLIST crossReference type CDATA #IMPLIED>

<!ELEMENT linkage (#PCDATA)>
<!ATTLIST linkage type CDATA #IMPLIED >
<!ATTLIST linkage code CDATA #IMPLIED >

<!ELEMENT relationship (#PCDATA | relatedDoc )* >
<!ATTLIST relationship type %relationshipTypes; >
<!ATTLIST relationship role %relationshipRoles; >

<!ELEMENT relatedDoc (#PCDATA | docid | title)* >
<!ATTLIST relatedDoc dtd CDATA #IMPLIED >
```

```
<!ELEMENT docimage (#PCDATA)>
<!ATTLIST docimage type CDATA #IMPLIED >
<!ATTLIST docimage href  CDATA #IMPLIED >
<!ATTLIST docimage language %languages; >

<!ELEMENT note (#PCDATA) >

<!-- End of the GLIN Document Type Definition -->
```

A sample coding for GLIN documents using the DTD we define above is as follows:

```
<glindoc>
pwd<docid>GLIN-99999</docid>
<country code="1">United States</country><class>Laws</class>
<number>Public Law 105-121</number>
<title>Export-Import Bank Reauthorization Act of 1997</title>
<subjectIndex type="controlled" thesaurus="glin" language="English">
<term id="182" order="1">Banks &amp; banking</term></subjectIndex>
<issuance><date month="11" day="26" year="1997">November  26, 1997</date></issuance>
<issue><publication>United States Statutes at Large</publication>
<number>111 Stat. 2528</number>
<date month="11" day="26" year="1997">November  26, 1997</date>
<specifics></specifics></issue>
<summary>Public Law 105-121 (111 Stat. 2528) of Nov. 26, 1997, the
Export-Import Bank Reauthorization Act of 1997 - Amends the
Export-Import Bank Act of 1945 to extend the authority of the
Export-Import Bank of the United States through FY 2001.  Reauthorizes
the Bank&amp;quot;s tied aid credit program. &amp;lt;p&amp;gt; (Sec. 4) Extends from FY 1997
through 2001 Bank authority to provide financing for the export of
nonlethal defense articles or services whose primary end use will be
for civilian purposes. &amp;lt;p&amp;gt; (Sec. 5) Revises Bank procedures governing
the denial of the extension of credit to foreign countries based on
the national interest to: (1) require the President to consult with
specified congressional committees before determining that such a
denial is in the U.S. national interest; and (2) require written
notification to the President of the Bank of such determination,
including the applications or categories of applications for credit
which should be denied. &amp;lt;p&amp;gt; (Sec. 6) Directs the General Counsel of
the Bank to designate an attorney to serve as Assistant General
```

Counsel for Administration, whose duties shall include oversight of and advice to Bank directors, officers, and employees on personnel and other administrative law matters. <p> (Sec. 7) Requires the Board of Directors of the Bank to: (1) take prompt measures to promote the expansion of its loan, guarantee, and insurance programs in sub-Saharan Africa; (2) establish an advisory committee to advise it on the implementation of policies and programs to support such expansion; and (3) report annually to the Congress on steps it has taken to implement such policies and programs and any advisory committee recommendations. <p> (Sec. 8) Revises the composition of the Advisory Committee of the Bank to include the appointment of not fewer than two members from the labor community. <p> (Sec. 9) Directs the President of the Bank to: (1) enhance the Bank"s capacity to provide information about its programs to small and rural companies which have not previously participated in them; and (2) report to the Congress on such activities within one year of enactment of this Act. <p> (Sec. 11) Includes child labor as a human rights criterion that could serve as the basis for a presidential determination that an application for Bank credit should be denied for nonfinancial or noncommercial considerations. <p> (Sec. 12) Requires the President, if the Russian military or Government has transferred an SS-N-22 missile system to China and such transfer represents a threat to U.S. security, to notify the Bank as soon as practicable. Directs the Bank Board of Directors to deny any guarantee, insurance, or extension of credit in connection with purchases of Russian goods or services if so directed by the President. <P>Contains 12 sections in 5 pages. </summary>

<linkage type="url-video" code="movie1.mpeg">Video clip of conference at U.S. House c
Representatives</linkage>
<relationship type="Regulates" role="active"><relatedDoc><docid>10620</docid>
<title>Bahrain -- Legislative Decree No.7 of 1991. Amends certain provisions of Law .
</relatedDoc></relationship><relationship type="Repealed by" role="passive">
<relatedDoc><docid>10620</docid>
<title>Bahrain -- Legislative Decree No.7 of 1991. Amends certain provisions of Law .
</relatedDoc></relationship>

</glindoc>

<glindoc>
<docid>GLIN-100900</docid><country code="1">United States</country>
<class>Other</class><number></number>

<title>Grizzly Bear Recovery Plan</title>
<subjectIndex type="controlled" thesaurus="glin" language="English"></subjectIndex>
<issuance><date month="09" day="10" year="1993">September 10, 1993</date></issuance>
<issue>
<publication>U.S. Fish and Wildlife Service. 1993. Grizzly bear recovery plan. Misso
MT 181 pp.</publication>
<number></number><date month="01" day="01" year="1993">January 01, 1993</date>
<specifics></specifics></issue>
<summary>The Grizzly Bear Recovery Plan was developed by the
U.S. Forest Service policy document under Section 3(f) of the
Endangered Species Act. Section 3(f) requires the Service to develop
recovery plans for the conservation of species listed as threatened
under the Act to the point where legal protection under the Act is no
longer necessary. The grizzly bear (Ursus arctos horribilis) was
listed as threatened on July 28, 1975. The original recovery plan was
approved on January 29, 1982, and the 1993 version is the first
revision of that plan. The recovery plan aims to allow the Forest
Service to delist of the grizzly bear by achieving recovery targets.
It proposes actions such as limiting habitat loss or degradation from
road building, timber harvest, oil and gas exploration and
development, mining, and recreation, improving knowledge of the
relationship between bear density and habitat type, and developing
techniques for moving bears successfully into areas where populations
are in need of augmentation.</summary>

<linkage type="url" code="http://www.cs.umbc.edu/~kalpakis/ELIS/northsp.gif">
See also an image of the Grizzly Bear habitat in the northwest.</linkage>
<relationship type="Implements" role="active">
<relatedDoc><docid>100910</docid><title>Endangered Species Act of 1973</title></relat
</relationship>

</glindoc>

<glindoc>
<docid>GLIN-100910</docid><country code="1">United States</country>
<class>Statutes</class><number></number>
<title>Endangered Species Act of 1973</title>
<subjectIndex type="controlled" thesaurus="glin" language="English"></subjectIndex>
<issuance><date month="12" day="28" year="1973">December 28, 1973</date></issuance>
<issue><publication>United States Code</publication>
<number>16 U.S.C. \247\247 1531-1544</number>

<date month="01" day="01" year="1998">January 01, 1998</date><specifics></specifics>
<summary>The Endangered Species Act (ESA) is the primary U.S. federal statute for protecting threatened and endangered species. An "endangered species" is any species that is in danger of extinction throughout all or a significant portion of its range. A "threatened species" is any species that is likely to become an endangered species within the forseeable future throughout all or a significant portion of its range. Distinct, vertebrate population segrments or subspecies may be listed separately. The U.S. Fish and Wildlife Service is responsible for listing all species except certain marine species that are listed by the National Marine Fisheries Service. Citizens can petition to list or change the status of a species. Listings must be made solely on the basis of the best available biological assessments. Critical habitat must be identified around the same time as a listing is made.</summary>

<linkage type="url" code="http://www.law.cornell.edu/uscode/16/1531.notes.html">
Framework law for the Grizzly Bear Recovery Plan</linkage>
<relationship type="Implemented by" role="passive"><relatedDoc>
<docid>100900</docid><title>Grizzly Bear Recovery Plan</title></relatedDoc></relation

</glindoc>

<glindoc>
<docid>GLIN-100920</docid><country code="1">United States</country>
<class>Statutes</class><number></number>
<title>National Environmental Policy Act of 1969</title>
<subjectIndex type="controlled" thesaurus="glin" language="English"></subjectIndex>
<issuance><date month="01" day="01" year="1970">January 01, 1970</date></issuance>
<issue><publication>United States Code</publication>
<number>42 U.S.C. \247\247 4321-4370(d)</number>
<date month="01" day="01" year="1998">January 01, 1998</date>
<specifics></specifics></issue>
<summary>The National Environmental Policy Act (NEPA) requires government agencies to prepare an environmental impact statement (EIS) for any major federal action that significantly affects the quality of the human environment. While only major federal actions trigger the EIS requirements, many "private" projects are reviewed pursuant to NEPA because they rely on federal financing, assistance, or project approval. Major federal actions include the adoption of

```
most official policies, formal plans, or programs, as well as the
approval of specific projects.

In addition to the EIS requirements, NEPA establishes and
coordinating long-term environmental research programs and
data clearinghouses.  Finally, it contains a number of
miscellaneous environmental policy provisions, including a
provision for financial assistance to citizen environmental
groups.</summary>

</glindoc>
```

Users can encode a plain text file to be a conformant XML GLIN document with ease by using any publicly or commercially available XML editor. Xemacs (http://www.xemacs.org) provides a nice XML editor for both Unix and Windows NT platforms. IBM, among other vendors, provides a reasonable public domain Java-based XML editor.

3.1.2 Relational Schema for Heterogeneous Legal Documents

In order to facilitate the integrative management and querying of legal documents of various structures (DTDs), we have designed a relational schema that enables the storage, querying, and retrieval of such legal documents. The schema of that database is as follows:

```
-- Schema for document database
-- have a table for metadata, the text of the doc,
--      sources, xml/xsl/css object names and associated URLs
--      as well as associated transforms with each DTD (XSLT and CSS)

CREATE TABLE doc_metadata_tb (
   docid       varchar2(32),  -- primary key,
   dtd_name    varchar2(64),  -- name of DTD for the metadata, if applicable
   title       varchar2(1024),
   docDate     date,          -- associtaed date for document (publication, approval,
   country_id  varchar2(8),   -- acronym for the country document originates from
   citation    varchar2(512), -- citation for the document
   metadata    LONG,          -- XML metadata in the specified DTD for the document
   sourceURL   varchar2(512), -- URL of location for the source of the document (spec
   source_id   int            -- identifier for the name of the source (agency, org.,
);
```

```
CREATE TABLE doc_text_tb (
   docid       varchar2(32),  -- primary key
   dtd_name    varchar2(64),  -- name of DTD for the full-text, if applicable
   fullText    LONG,
   type        varchar2(10),  -- type of external file referenced by extHREF
   extHREF     varchar2(512), -- URL to file of full-text (external)
   sourceURL   varchar2(512), -- URL of location for the full-text of the document
   source_id   int            -- identifier for the name of the source (agency, org.,
);

CREATE TABLE doc_history_tb (
   docid       varchar2(32),
   actionDate  date,
   person      varchar2(64),  -- person acted on document
   actionType  varchar2(10)   -- type of action taken (insert, delete, update,etc)
);

CREATE TABLE doc_sources_tb (
   source_id   int,            -- primary key
   source_name varchar2(256),  -- name of source agency, organization, etc
   baseURL     varchar2(512)   -- URL to home Web page for organization
);

CREATE TABLE xml_transforms_tb (
   detailLevel int,            -- level of detail of output (1:succint to 10:complete
   dtd_name    varchar2(64),   -- name of DTD on which to apply XSLT transformation
   xsl_sheet   varchar2(64),   -- name of XSLT stylesheet for the transformation
   css_sheet   varchar2(64)    -- name of CSS stylesheet for the formating after tran
);

CREATE TABLE xml_objects_tb (
   name  varchar2(64),         -- name of DTD/XSLT/CSS object
   url   varchar2(1024)        -- URL for locating the object
);
```

3.1.3 XSL for GLIN Document Presentation

In order to allow the users of the GLIN system to view the GLIN documents in a familiar and intuitive manner, we develop an XSL stylesheet for processing GLIN documents encoded according to the DTD above. XSL stylesheets are used to format a GLIN document

encoded in XML at run–time. Using XSL stylesheets, we can customize presentation of documents to user's preferences at run–time with minimal maintaince cost. In fact, we can allow sophisticated users to define their own presentation format for GLIN documents.

The default XSL stylesheet for the GLIN documents that is used by our prototype is as follows:

```
<xsl:stylesheet xmlns:xsl='http://www.w3.org/1999/XSL/Transform' version='1.0'>

<xsl:template match='glindoc'>

<table width='95%' border='0'><tr><td>

    <table border='1' cellpadding='10' cellspacing='0' width='100%' bordercolor='#F8C
    <tr><td class='notecard' bgcolor="white">
<!-- xsl:apply-templates select='docid' -->
<xsl:apply-templates select='country'/>
<xsl:apply-templates select='class'/>
    </td></tr>
    </table>

<xsl:apply-templates select='title'/>

<br></br>

        <xsl:apply-templates select='subjectIndex'/>
        <xsl:apply-templates select='issueance'/>
        <xsl:apply-templates select='issue'/>

<br></br>

<xsl:call-template name='doCitations'/>
<xsl:call-template name='doCrossReference'/>
<xsl:call-template name='doRelationship'/>
<xsl:call-template name='doDocimage'/>

        <xsl:apply-templates select='summary'/>

    </td></tr></table>

</xsl:template>
```

```
<xsl:template name='doCitations'>
        <xsl:if test='count(citation) > 0'>
          <DIV>
          <B>Citation(s)</B>
          <UL>
          <xsl:for-each select='citation'>
          <xsl:sort select='.'/>
           <xsl:apply-templates select='.'/>
          </xsl:for-each>
          </UL>
          </DIV>
        </xsl:if>
</xsl:template>

<xsl:template name='doCrossReference'>
        <xsl:if test='count(crossReference) > 0'>
          <DIV>
          <B>Cross-reference(s)</B>
          <UL>
          <xsl:for-each select='crossReference'>
          <xsl:sort select='.'/>
           <xsl:apply-templates select='.'/>
          </xsl:for-each>
          </UL>
          </DIV>
        </xsl:if>
</xsl:template>


<xsl:template name='doRelationship'>
        <xsl:if test='count(relationship)>0 '>
  <DIV>
          <B>Relationship(s)</B>
          <UL>
  <xsl:for-each select='relationship'>
    <xsl:sort select='@type'/>
            <xsl:apply-templates select='.'/>
  </xsl:for-each>
          </UL>
          </DIV>
        </xsl:if>
```

```
</xsl:template>

<xsl:template name='doDocimage'>
        <xsl:if test='count(docimage)>0 '>
  <DIV>
          <B>See also</B>
          <UL>
  <xsl:for-each select='docimage'>
    <xsl:sort select='.'/>
            <xsl:apply-templates select='.'/>
  </xsl:for-each>
          </UL>
          </DIV>
        </xsl:if>
</xsl:template>


<xsl:template match='docid'>
<div>
   <b><i>Document ID: </i></b>
   <xsl:value-of select='text()'/>
</div>
</xsl:template>

<xsl:template match='country'>
<div>
  <B>Country: </B> <xsl:value-of select='text()'/>
</div>
</xsl:template>

<xsl:template match='class'>
<div>
  <b>Class: </b> <xsl:value-of select='text()'/>
</div>
</xsl:template>

<xsl:template match='title'>
<div>
  <center>
  <table border='0' width='80%' cellspacing='0' cellpadding='5'>
  <tr><th align='left'>
  <B>Title </B>
```

```
  </th></tr></table>
  <table border='1' width='80%' cellspacing='0' cellpadding='5'>
  <tr><td class='title' bgcolor='white' bordercolor='#F8CE0A'>
  <xsl:if test='string-length(ancestor::glindoc/number)>0'>
     <xsl:value-of select='ancestor::glindoc/number'/>:
  </xsl:if>
  <xsl:value-of select='text()'/>
  </td></tr></table>
  </center>
</div>
</xsl:template>

<xsl:template match='subjectIndex'>
  <xsl:if test='count(term) > 0 '>
    <div>
    <b>Subject Index</b>
    <ul>
    <xsl:for-each select='term'>
      <xsl:sort select='.'/>
<LI><xsl:apply-templates select='.'/></LI>
    </xsl:for-each>
    </ul>
    </div>
  </xsl:if>
</xsl:template>


<xsl:template match='issueance'>
    <div>
    <b>Issuance: </b>
    <xsl:value-of select='date'/>
    </div>
</xsl:template>

<xsl:template match='date'>
 <B>Date: </B>
 <xsl:value-of select='text()'/>
</xsl:template>

<xsl:template match='issue'>
<div>
  <B>Issue: </B>
```

```
  <xsl:value-of select='number'/>
  <xsl:text>. (</xsl:text>
  <i><xsl:value-of select='publication'/></i>
  <xsl:text>, </xsl:text>
  <xsl:value-of select='date'/>
  <xsl:text>)</xsl:text>
</div>
</xsl:template>

<xsl:template match='citation'>
  <LI>
  <xsl:value-of select='text()'/>
  [<xsl:value-of select='@type'/>]
  </LI>
</xsl:template>


<xsl:template match='crossReference'>
  <LI>
    <xsl:value-of select='text()'/>
    <xsl:if test='count(title)> 0'>
    <DIV>
    Title(s)
    <UL>
    <xsl:for-each select='title'>
      <xsl:sort select='.'/>
      <LI>
<xsl:if test='string-length(text())>0'>
<I><xsl:value-of select='text()'/></I>
        </xsl:if>
      </LI>
    </xsl:for-each>
    </UL>
    </DIV>
    </xsl:if>

    <xsl:if test='count(linkage)> 0'>
    <DIV>
    Linkage(s)
    <UL>
    <xsl:for-each select='linkage'>
      <xsl:sort select='.'/>
```

```
      <LI>
      <A HREF='{@code}'><xsl:value-of select='text()'/></A>
      </LI>
    </xsl:for-each>
    </UL>
    </DIV>
    </xsl:if>
  </LI>
</xsl:template>


<xsl:template match='relationship'>
  <LI>
    <xsl:value-of select='@type'/>
    <xsl:value-of select='text()'/>
    <xsl:if test='count(relatedDoc)>0 '>
      <UL>
      <xsl:for-each select='relatedDoc'>
        <xsl:sort select='.'/>
        <LI>
        <A HREF='http://bazak:9080/getDoc,jsp?docid={docid}'>
        <xsl:value-of select='title'/>
        <xsl:value-of select='text()'/>
        </A>
        </LI>
      </xsl:for-each>
      </UL>
    </xsl:if>
  </LI>
</xsl:template>


<xsl:template match='docimage'>
  <LI>
    <A HREF='{@href}'>
        <xsl:value-of select='text()'/>
    </A>
    <xsl:text> </xsl:text>
    <xsl:if test='string-length(@type)>0'>
       [<xsl:value-of select='@type'/>]
    </xsl:if>
  </LI>
```

```
</xsl:template>


<xsl:template match='summary'>
<xsl:if test='string-length(text())>0'>
<div>
<b>Summary</b>
<BR>
<DIV>
<xsl:apply-templates/>
</DIV>
</BR>
</div>
</xsl:if>
</xsl:template>

<xsl:template match='linkage'>
    <A HREF='{@code}'>
    <xsl:value-of select='text()'/>
    </A>
</xsl:template>



</xsl:stylesheet>
```

3.1.4 Indexing

The prototype we developed allows for sophisticated search queries that include not only traditional relational queries but also full-text queries. This is accomplished by building appropriate text indexes in Oracle 8.1.5. In order to enable full-text searches within sections (elements) of the XML-formatted GLIN documents, we first create appropriate tags and preferences for Oracle 8.1.5:

```
--
--
-- Create Table with the element names of the DTDs that are seachable
--

CREATE table searchable_tags_tb (
   dtd_name     varchar2(64),
   dtd_tag      varchar2(128)
);

INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'glindoc');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'country');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'class');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'number');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'title');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'subjectIndex');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'issueance');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'issue');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'date');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'summary');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'citation');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'crossReference');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'linkage');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'relationship');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'publication');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'specifics');
INSERT INTO searchable_tags_tb VALUES('glindoc_v1.dtd', 'relatedDoc');



-------------------------------------------------------------------------
-- Create Preferences and Tables for Gists, Themes, and Markups

-------------------------------------------------------------------------
-- Create the preferences for the datastore and document filter
--
CALL CTX_DDL.CREATE_PREFERENCE('lawDataStore', 'DIRECT_DATASTORE');
CALL CTX_DDL.CREATE_PREFERENCE('lawFilter', 'NULL_FILTER');
-------------------------------------------------------------------------


-------------------------------------------------------------------------
```

```
-- Create the preference for the Lexer
--
CALL CTX_DDL.CREATE_PREFERENCE('lawLexer', 'BASIC_LEXER');
CALL CTX_DDL.SET_ATTRIBUTE('lawLexer', 'INDEX_TEXT', 'YES');
CALL CTX_DDL.SET_ATTRIBUTE('lawLexer', 'INDEX_THEMES', 'YES');
--
-------------------------------------------------------------------------


-------------------------------------------------------------------------
-- Create the preference for the stemmer and fuzzy matcher
--
CALL CTX_DDL.CREATE_PREFERENCE('lawWordList', 'BASIC_WORDLIST');
CALL CTX_DDL.SET_ATTRIBUTE('lawWordList', 'Stemmer', 'ENGLISH');
CALL CTX_DDL.SET_ATTRIBUTE('lawWordList', 'Fuzzy_Match', 'ENGLISH');
--
-------------------------------------------------------------------------


-------------------------------------------------------------------------
-- Create Stoplist and initialize it to use all the stopwords from the
-- default stoplist
--
CALL CTX_DDL.CREATE_STOPLIST('lawStopList');
DECLARE
  CURSOR cursor_spw IS
       SELECT SPW_WORD
       FROM CTX_STOPWORDS
       WHERE SPW_STOPLIST = 'DEFAULT_STOPLIST' AND SPW_TYPE ='STOP_WORD';
  cursor_val cursor_spw%ROWTYPE;
BEGIN
  FOR cursor_val IN cursor_spw
    LOOP
      CTX_DDL.ADD_STOPWORD('lawStopList', cursor_val.SPW_WORD);
    END LOOP;
END;
.
RUN;
--
-- Add/Removestopwords and/or stopthemes
--
CALL CTX_DDL.ADD_STOPWORD('lawStopList', 'JSP');
CALL CTX_DDL.ADD_STOPTHEME('lawStopList', 'Kalpakis');
--CALL CTX_DDL.REMOVE_STOPWORD('lawStopList', 'XML');
```

```
--CALL CTX_DDL.REMOVE_STOPTHEME('lawStopList', 'Kalpakis');
--
--
-------------------------------------------------------------------------


-------------------------------------------------------------------------
-- Create Section Group
--
CALL CTX_DDL.CREATE_SECTION_GROUP('lawSections', 'XML_SECTION_GROUP');
--
-- Add/Remove section zones to the section group
--
DECLARE
  CURSOR cursor_spw IS
       SELECT DISTINCT dtd_tag
       FROM searchable_tags_tb;
  cursor_val cursor_spw%ROWTYPE;
BEGIN
  FOR cursor_val IN cursor_spw
    LOOP
      CTX_DDL.ADD_ZONE_SECTION('lawSections', cursor_val.DTD_TAG, cursor_val.DTD_TAG)
    END LOOP;
END;
.
RUN;
--
--CALL CTX_DDL.ADD_ZONE_SECTION('lawSections', 'DOC', 'DOC');
--CALL CTX_DDL.ADD_ZONE_SECTION('lawSections', 'TITLE', 'TITLE');
--CALL CTX_DDL.ADD_ZONE_SECTION('lawSections', 'AUTHOR', 'AUTHOR');
--CALL CTX_DDL.ADD_ZONE_SECTION('lawSections', 'SUMMARY', 'SUMMARY');
--
--CALL CTX_DDL.REMOVE_SECTION('lawSections', 'SUMMARY');
--
-------------------------------------------------------------------------


-------------------------------------------------------------------------
-- Create Gist, Theme, and Markup tables
--
CREATE TABLE docGist(query_id NUMBER, pov VARCHAR2(80), gist CLOB);
--
CREATE TABLE docTheme(query_id NUMBER, theme VARCHAR2(2000), weight NUMBER);
```

```
--
CREATE TABLE docMarkup(query_id NUMBER, document CLOB);
--
-------------------------------------------------------------------------

-------------------------------------------------------------------------
-- Create Indexes
--

CREATE INDEX doc_meta1_idx ON doc_metadata_tb(title) INDEXTYPE IS CTXSYS.CONTEXT
PARAMETERS('datastore lawDataStore
            lexer     lawLexer
            wordlist  lawWordList
            stoplist  lawStopList
            section group lawSections
            memory 20M');

CREATE INDEX doc_meta2_idx ON doc_metadata_tb(metadata) INDEXTYPE IS CTXSYS.CONTEXT
PARAMETERS('datastore lawDataStore
            lexer     lawLexer
            wordlist  lawWordList
            stoplist  lawStopList
            section group lawSections
            memory 20M');

CREATE INDEX doc_text1_idx ON doc_text_tb(fullText) INDEXTYPE IS CTXSYS.CONTEXT
PARAMETERS('datastore lawDataStore
            lexer     lawLexer
            wordlist  lawWordList
            stoplist  lawStopList
            section group lawSections
            memory 20M');

-------------------------------------------------------------------------
```

Note that Oracle 8.1.5 has certain limitations on text–indexing. For example, indexing a text column using user–defined section tags leads into non–availability of summarization (gist), thematic labeling (themes) and highlighting of text in the column. Further, let us mention that we have setup appropriate text–indexes for English text files in a variety of formats including PDF, Microsoft Word, Rich Text, Microsoft Excell, Wordperfect, etc. However, due to non-availability of appropriate data for GLIN, we do not demonstrate this

capability here.

3.1.5 Query Services

Users can compose queries using the query operators supported by Oracle 8.1.5. These include simple full-text search, sectioned full-text search, soundex and fuzzy queries, etc. Users can navigate the query results. They can also request that the system generate a summary of a GLIN document on–the–fly (e.g. a gist of the document), request that the system generate a list of thematic keywords that characterize a GLIN document (based on the knowledge–base available in Oracle 8.1.5), and also request a summary of a GLIN document with respect to a particular thematic keyword (e.g. thematic document gist). See Figures 4–10 for screen snapshots of the prototype system.

3.1.6 Security and Authentication

Secure access and authentication is implemented using the facilities provided by the Web server. Our prototype does not implement any additional mechanims. Nevertheless, given the architecture of the prototype, it is straightforward to incorporate a digital signature module to enable users to authenticate the contents of the GLIN documents they receive. Access control to the GLIN database is implemented by access control lists maintained as configuration files for the Web Server. The prototype implements a simple access control policy that as supported by the Web server.

3.1.7 Online GLIN Document Maintainance

Contributors and associates of the GLIN system are provided with two Web–based methods to insert, update, or delete documents from the GLIN document collection. One method is by working on the normalized database tables that are used for storing the GLIN documents. The other method is by working on the database table that contains the XML version of GLIN documents. In either case, contributors interact with the system only via a simple Web–based interface and they do not need any specialized software (except for a Web browser such as IE 5.0 or Netscape 4.7). Further, contributors can upload document files (e.g. ASCII, PDF, XML, Microsoft Word, or any other format) to the server via the Web. The prototype can index documents in a variety of formats, including PDF, Microsoft Word, RTF, etc. See Figures 11, 12, 13, and 14 for screen snapshots of the prototype system.

Appendix A: Summary of Activities for 1996–1997

CESDIS has been collaborating, since 1995, with NASA and the U.S. Library of Congress in the development of the Global Legal Information Network (GLIN) (http://glin.gsfc.nasa.gov). Since June 1996, I have been the technical leader of the GLIN project at CESDIS. Further, except for the summer months of 1996, in which I had 2 summer students (a high–school student and a sophomore college student), I was the only programmer available to this project.

GLIN is an on-line repository of legal instruments, providing global access to the legal information of participating nations. Currently there are 22 nations involved in the project. The primary goal of the GLIN system is to provide efficient, flexible, and reliable access to authentic, accurate, and current legal information. The GLIN member nations have committed to provide the appropriate content to its legal digital library.

Efforts on GLIN are on two fronts/phases running in parallel. Phase 1, the upgrade phase, calls for upgrading the Law Library's prototype with additional functionality required by the Law Library's staff. Phase 1 has three tasks. Task 1 is the upgrading of the prototype, Task 2 is surveying infrastructure of member countries, and Task 3 analyzing the communications system. Phase 2, the enhancement phase, calls for designing and developing the next generation system.

I developed (2) prototypes for Phase 1. The first one was using the INQUERY text-retrieval engine available at the Law Library. This prototype was eventually abandoned, primarily because of various limitations imposed by using INQUERY. The second one is using a WAIS text-retrieval engine based on the vector-space model for text retrieval. This prototype is the current GLIN prototype available at the http://glin.gsfc.nasa.gov URL. This prototype was first released at the end of January of 1997. The current version of this prototype was released in March 1997. I completed all the sub-tasks of Task 1 in Phase 1 that were requested except for handling multi-lingual legal instruments in their native character set/language. This subtask of Task 1, Phase 1, had been moved to Phase 2, since completing it would had required too drastic changes to the existing system. Furthermore, I incorporated certain additional features in this prototype that are envisioned in Phase 2. The rationale is to demonstrate preliminary versions of them to the GLIN user community at an early stage in order to capture their requirements for delivering a successful system. The current version of the GLIN prototype serves as the bridge between Phase 1 and 2. The current GLIN prototype consists of a database (Postgres) server, a WAIS server, a Web server, together with application software built using the functionality provided by the database and WAIS servers, and interacting with the users primarily via the Web server. Currently, legal documents are submitted to the data servers in SGML-format, and accessed either via SQL or Z39.50-type queries. Documents are indexed using the Legal Thesaurus developed by the U.S. Library of Congress, and their full-text summary (currently mostly English). In addition, digitized images of the legal documents are also stored in the system.

At the same time, I have been working on the architecture of the GLIN system in Phase 2. The architecture for GLIN is based on the agent–oriented programming approach and is inspired by ARPA's reference architecture for the intelligent integration of information. I also collaborated with Dr. Susan Hoban on refining the GLIN Project Plan.

I have demonstrated the use of ACTS communications capabilities for GLIN at the ADL'96 Conference. This demonstration was made possible by the assistance of technical staff members of Code 930, and especially Mr. Pat Garry.

I have presented my agent-based architecture for the next generation GLIN system at the 3rd Annual GLIN Directors meeting held at the Library of Congress in September 1997.

I have also given a tutorial during the February 1997 GLIN Training Session at the Library of Congress on CGI and Javascript scripts.

I have written the GLIN/Digital Libraries and Electronic Commerce attachments to the proposal submitted by SCDC (Code 930) and CESDIS to the US-Israel technology commission in the Fall of 1996.

I coauthored a paper with title "The Global Legal Information Network (GLIN)" which appeared in The American University Law Review, Vol. 46, No. 2, pp. 477–491, December 1996.

I was member of the Working Group on Digital Libraries and Electronic Commerce at the ACM Workshop on Strategic Directions in Computing Research held at MIT in July 1997. I coauthored a paper, based on the conclusions of that Working Group, with title "Electronic Commerce and Digital Libraries: Towards a Digital Agora", which appeared in ACM Computing Surveys, Vol. 28, No. 4, pp. 818–835, December 1996.

I completed a technical paper, coauthored with Bella Bellagradek (Ph.D. student) and Yelena Yesha, on "Strategies for Maximizing Seller's Profit under Unknown Buyer's Utility Values" which I submitted to the CASCON'97 Conference organized by IBM and the NRC, Canada. Suppose there is a seller that has an unlimited number of units of a single product for sale. The seller at each moment of time posts a price for his/her product. Based of the posted price, at each moment of time, a buyer decides whether or not to buy a unit of that product from the seller. The only information about the buyer to the seller is the seller's sales history. Further, I assume that the maximal unit price the buyer is willing to pay does not change over time. The question then is how should the seller price his/her product to maximize profits? To address this question, I use the notion of loss functions. Intuitively, a loss function is a measure, at each moment of time, of the lost opportunity to make a profit. In particular, I provided a polynomial–time algorithm that finds a pricing algorithm (strategy) for the seller that minimizes the cumulative (total) losses over time. Further, I presented preliminary results on pricing strategies that minimize the maximum possible loss at every moment of time. I also showed that there is no strategy minimizing both the total loss and the maximum loss at the same time.

I and an M.S. student of mine, George Durham, developed a multi–level security model for object–oriented databases. A technical paper based on this work will be presented at the 20th National Information Systems Security Conference on October 1997. Our model is based on, and extends the requirements of the Department of Defense 5200.28-STD, DOD Trusted Computer System Evaluation Criteria (TCSEC) dated December 1985, commonly known as the Orange Book. Currently, there does not exist a database model in any technology which meets the requirements of the Orange Book. There has been little interest outside of the U.S. Government and the academic community because the Orange Book is believed to focus on military needs rather than commercial interests. This is an unfortunate belief because in fact, commercial espionage is growing daily, and without proper protection, commercial information will be pilfered both nationally and internationally. Previous work has focused on Discretionary Access Controls (DAC), Mandatory Access Controls (MAC), or other security requirements not included in the Orange Book, but no work includes all three. We developed policies for access controls, inference controls, and implementation strategy based on the MAC, DAC, and other security requirements. The access authorization mechanism is based on a combination of DAC and MAC requirements, and the proposed model is easily extended to include other access requirements. We also described a system implementation.

Appendix B: Summary of Activities for 1998

My primary focus during this period was development work for GLIN. I undertook a number of developed activities related to the GLIN project. My efforts during this period were on developing a sequence of prototypes, experimenting with various approaches. The main line of the approach was to utilize the services provided by traditional relational database management systems in order to develop a GLIN prototype that addressed the requirements by the Law Library of the Library of Congress. The initial approach to use Postgress and Inquery, though shown that it can be done through a prototype, had certain drawbacks, that lead me to discard that approach. The next approach was to opt for DB2 or Oracle 8. I experimented with both of them, and both were shown to be appropriate. Based on the desire of the Law Library, the Oracle platform was selected. I developed a prototype system based on Oracle 8, using a combination of Java and Javascript to develop the various modules needed, while using the JDBC protocol to communicate with the database. The option to use the PL/SQL and Javascript was not selected, though quite appealing, since that would have lead into making the prospect of migrating into a non-Oracle platform infeasible. At this point, a prototype is running on Windows NT and Solaris platforms, as a Java application. Even though I was targeting that the prototype could also be used through the Web on standard Web browsers, due to limitations of the Netscape and Explorer browsers, currently, only a limited set of functions are fully available. I am exploring ways to get around those issues. A version of the prototype was demonstrated in the GSFC Technology Showcase in March 1998. Besides the development/prototyping work which was the main thrust of my effort for that period, various experiments were performed on bilingual text storage and retrieval, indexing and retrieval processing times, and capacity estimation. However, these efforts have not completed in this period.

As an extension to the basic GLIN prototype, in cooperation with colleagues from CESDIS, NASA GSFC, the Law Library, and the American University, we submitted a proposal, in response to CAN-97-05, with title "Integrating Legal and Environmental Information Systems" to the NASA's MTPE program. This proposal was selected for funding in the Spring of 1998, and is currently under way.

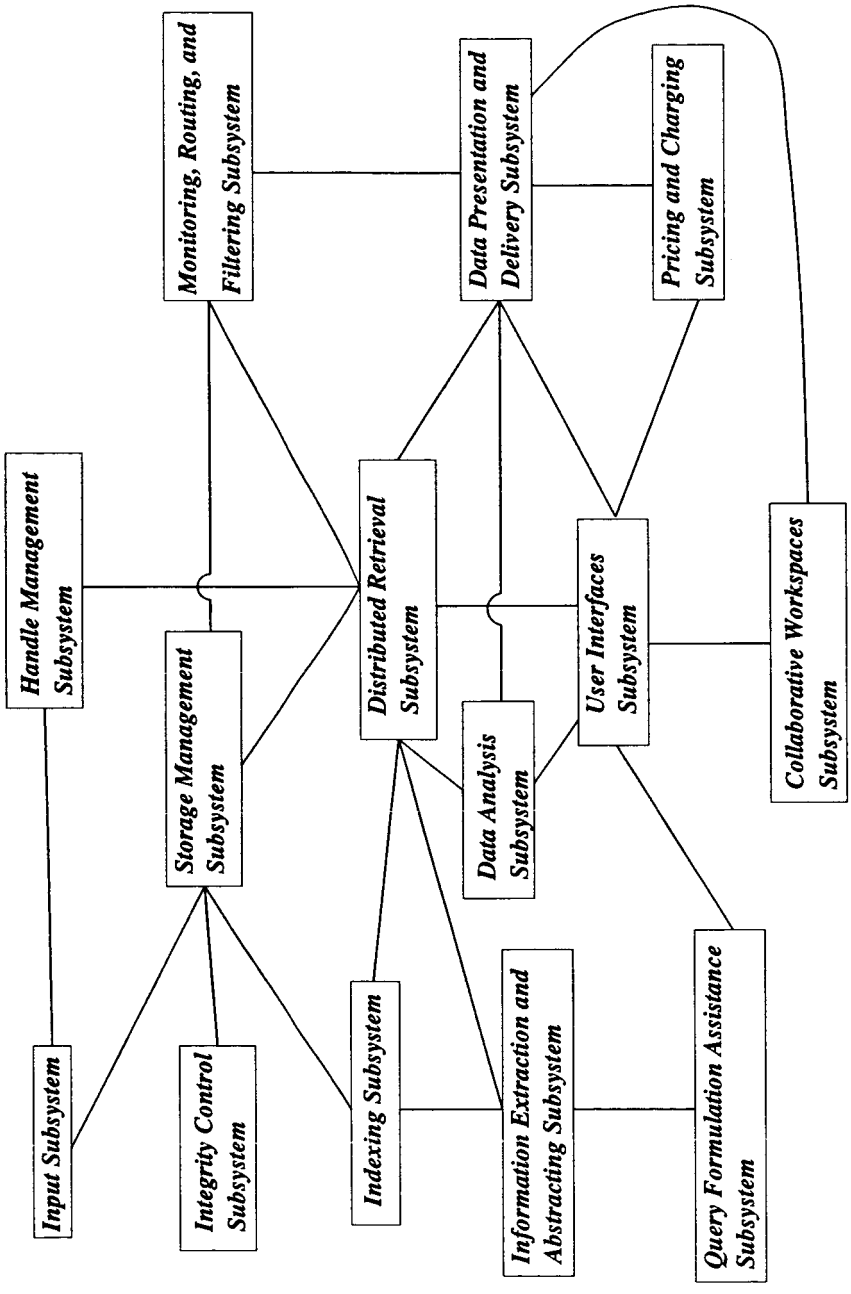

Figure 2: Subsystems in the GLIN System Architecture.

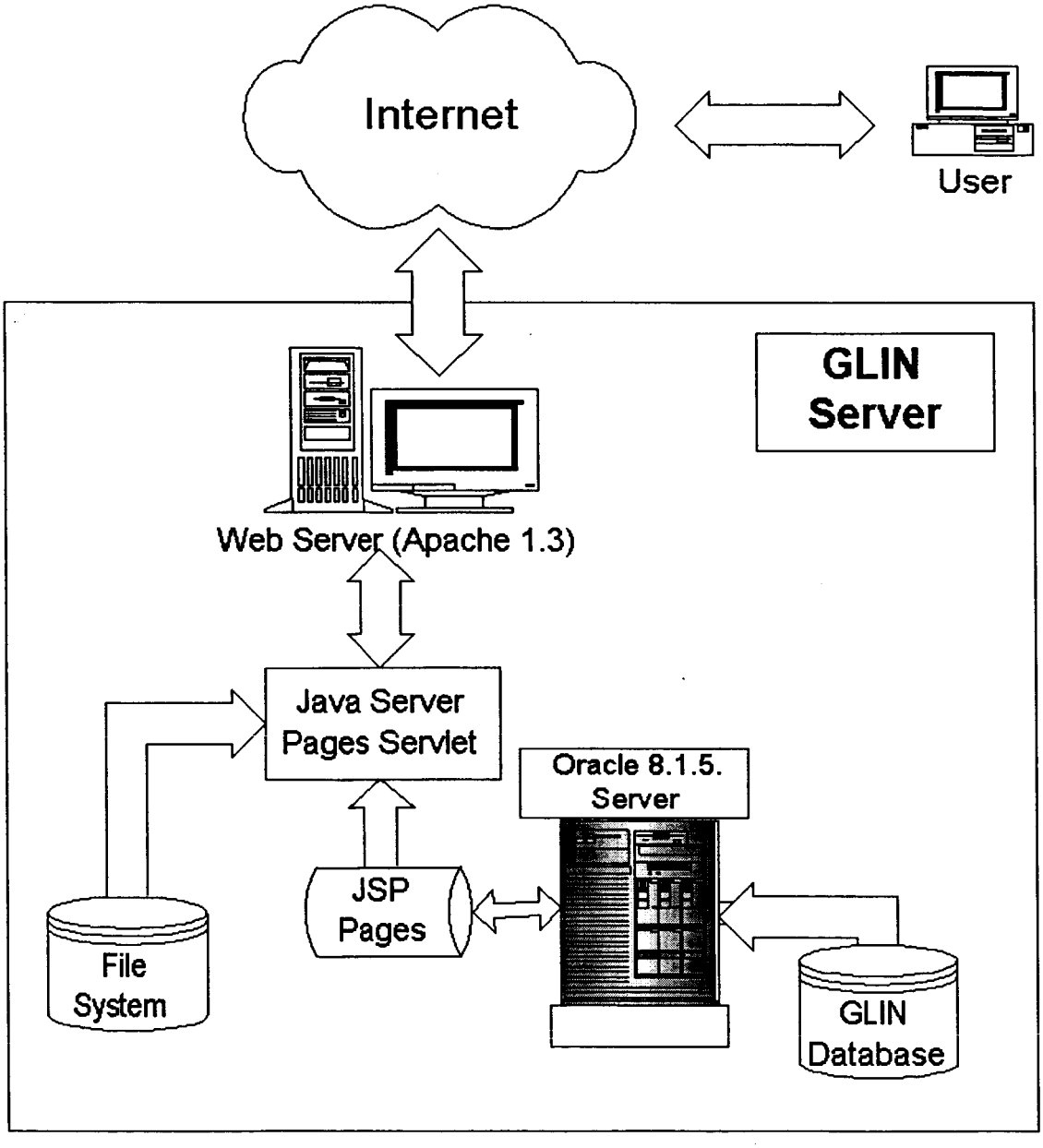

Figure 3: GLIN Prototype System.

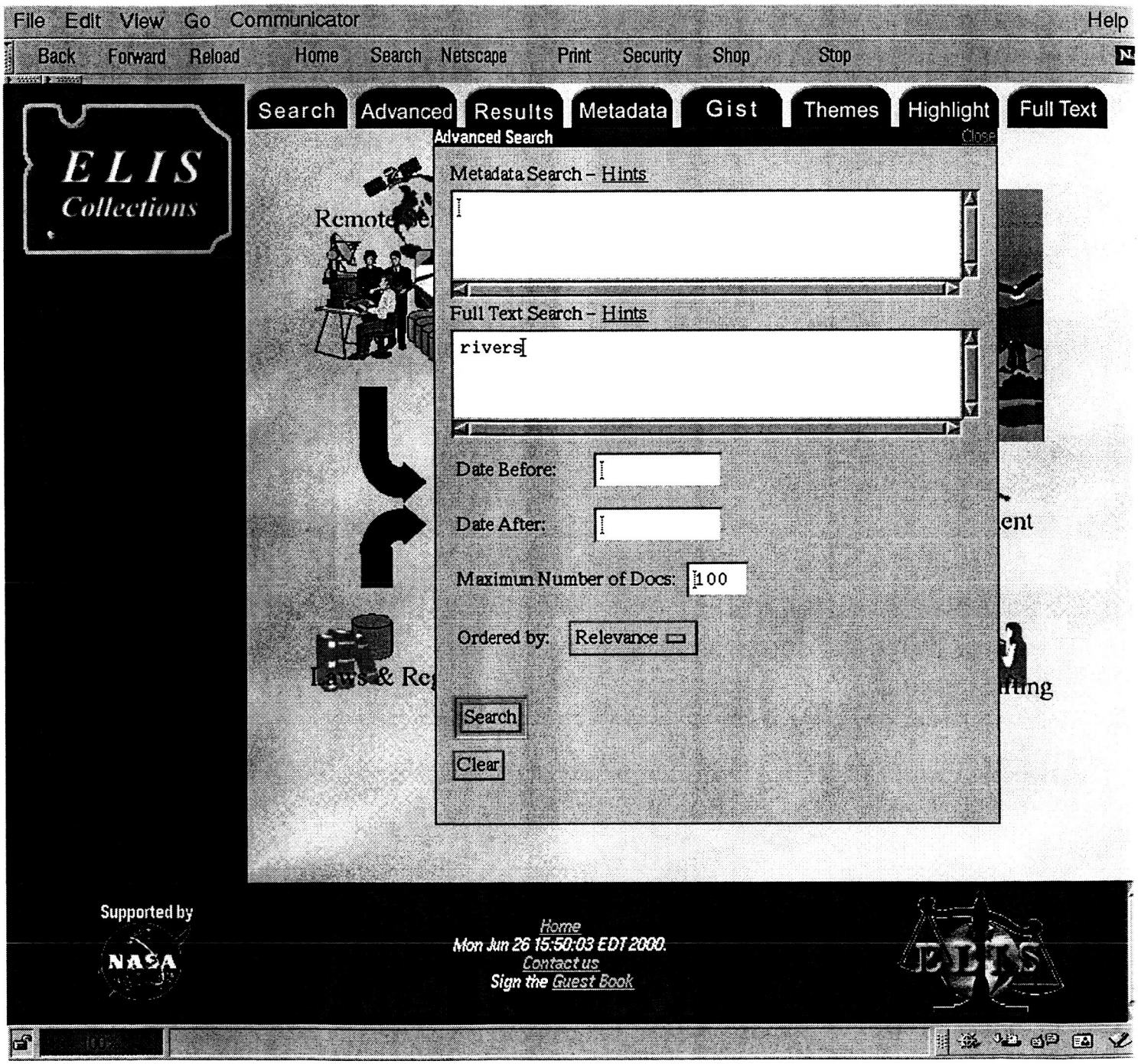

Figure 4: User specifies query to search for GLIN documents related to "rivers".

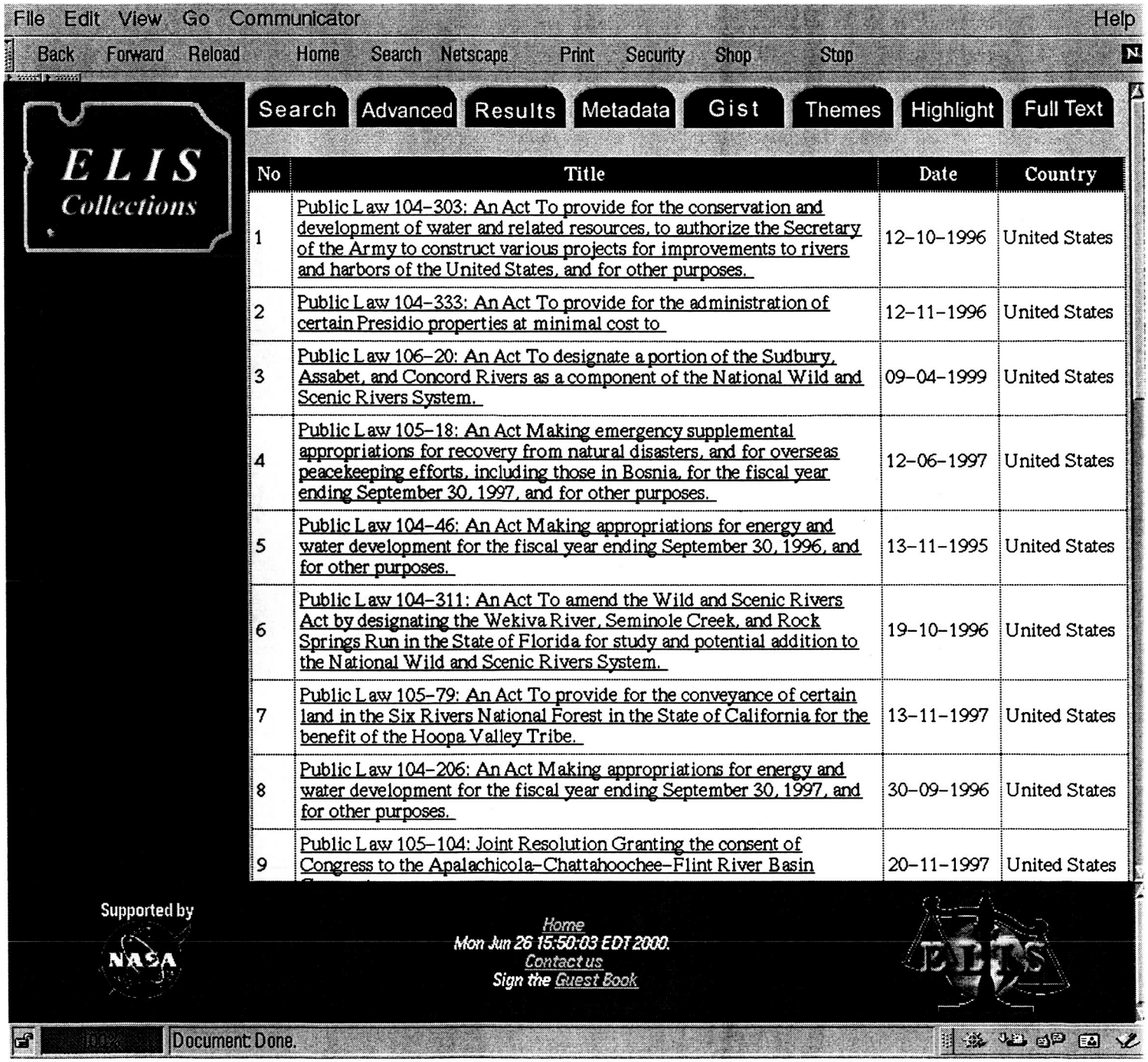

Figure 5: Results of user query.

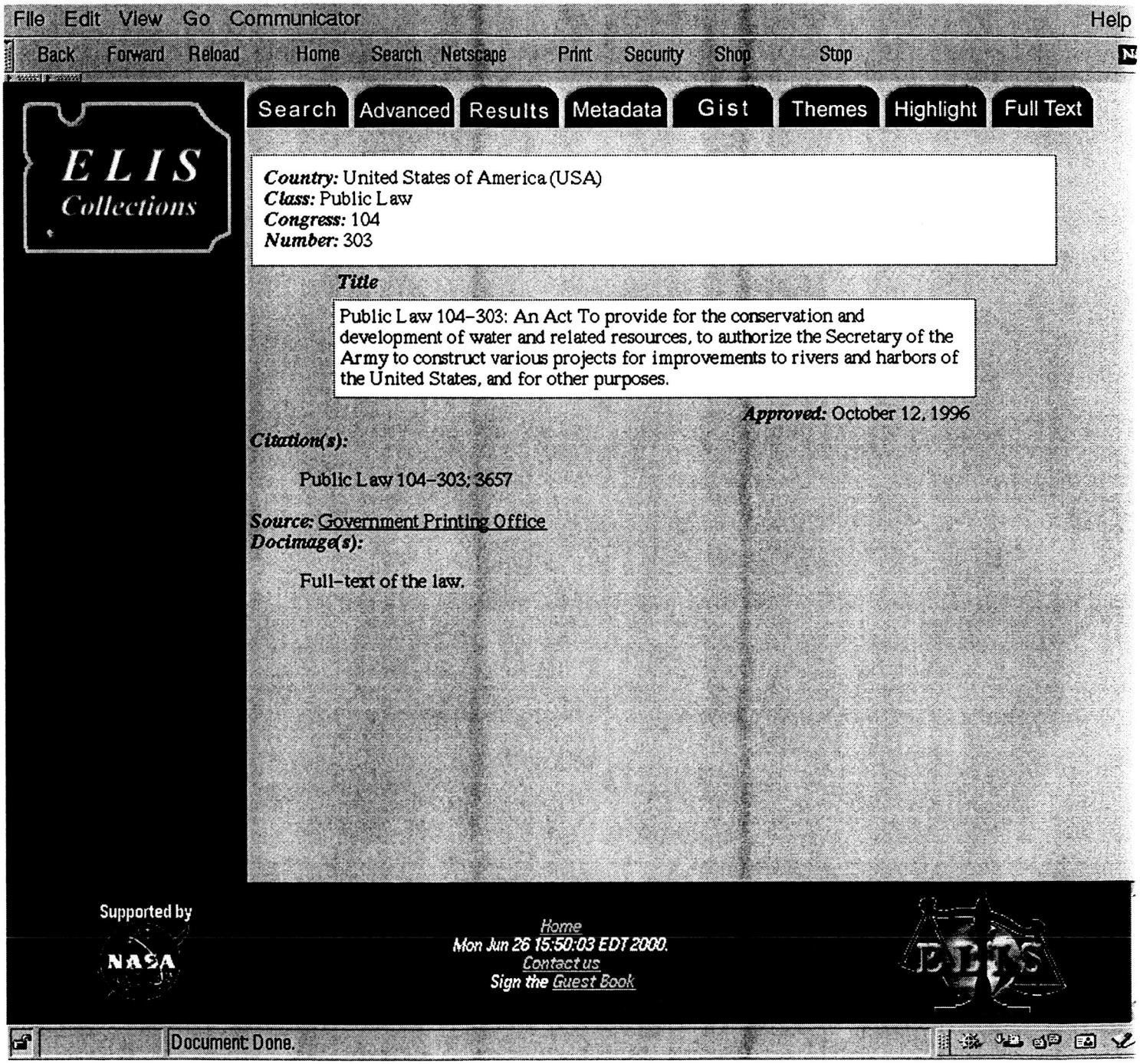

Figure 6: User requests to see metadata for the 1st hit.

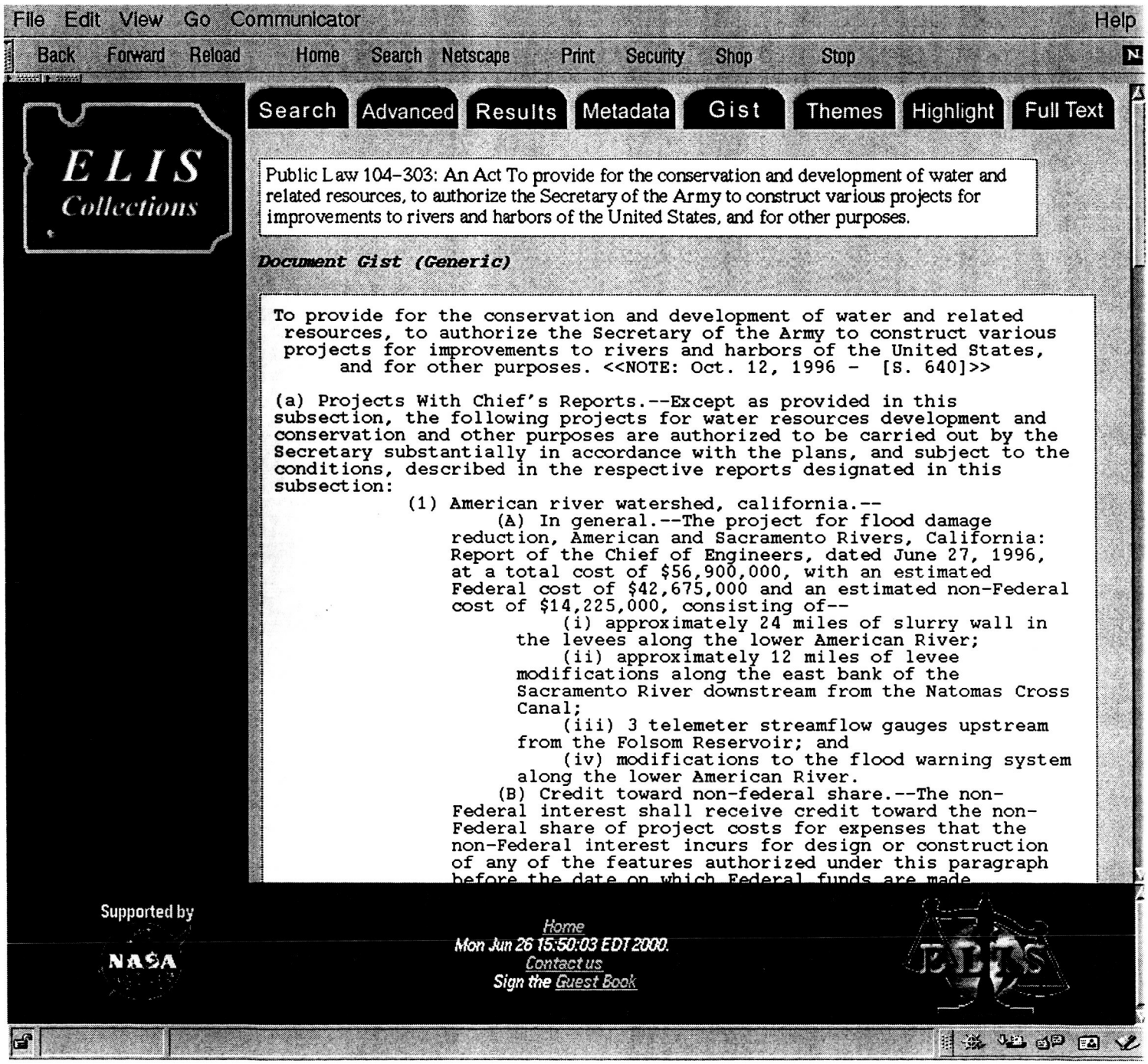

Figure 7: User requests to see the generic gist of the 1st hit.

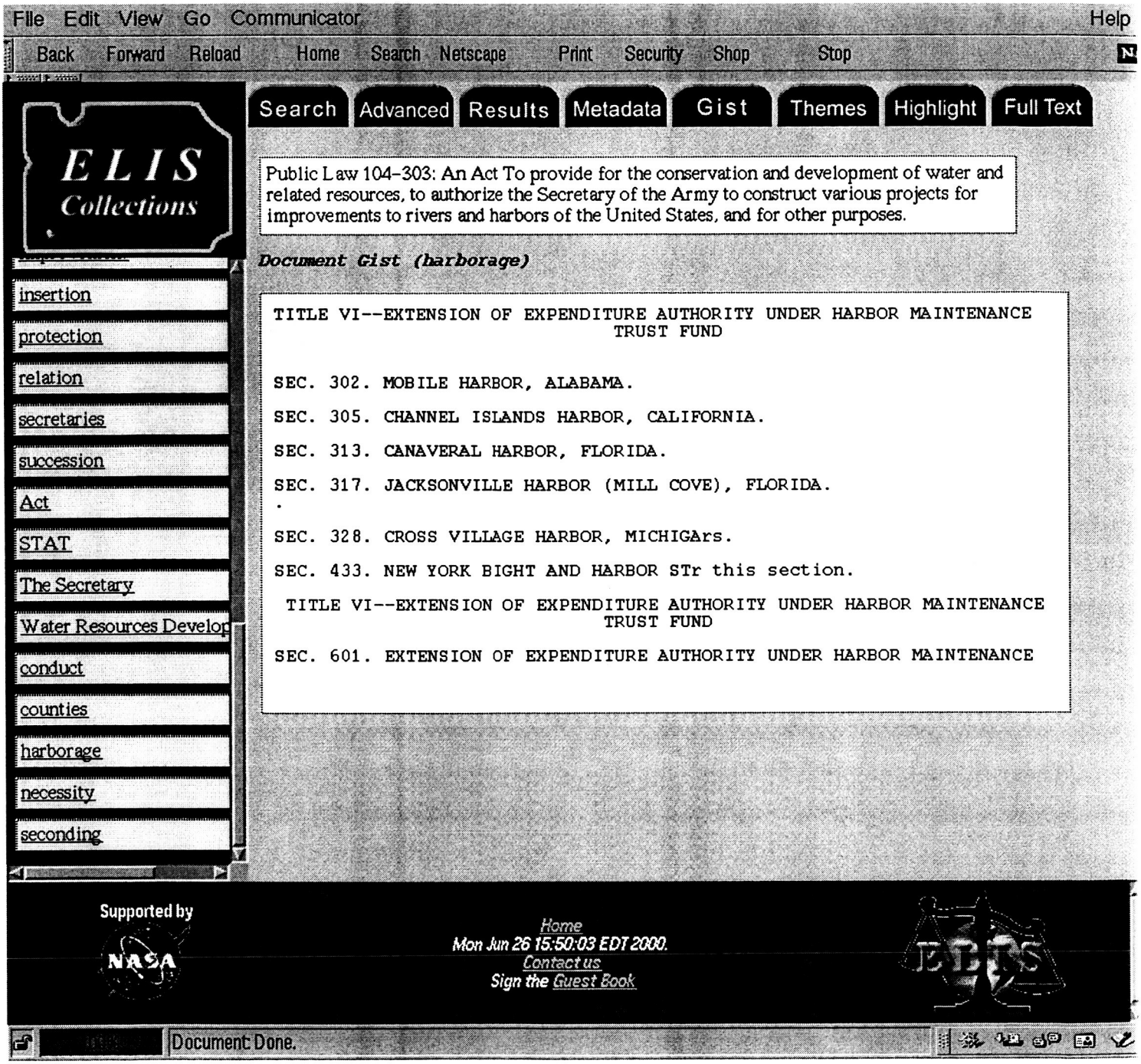

Figure 8: User requests a thematic list to be generated and then requests the gist of the 1st hit from the "harborage" point of view.

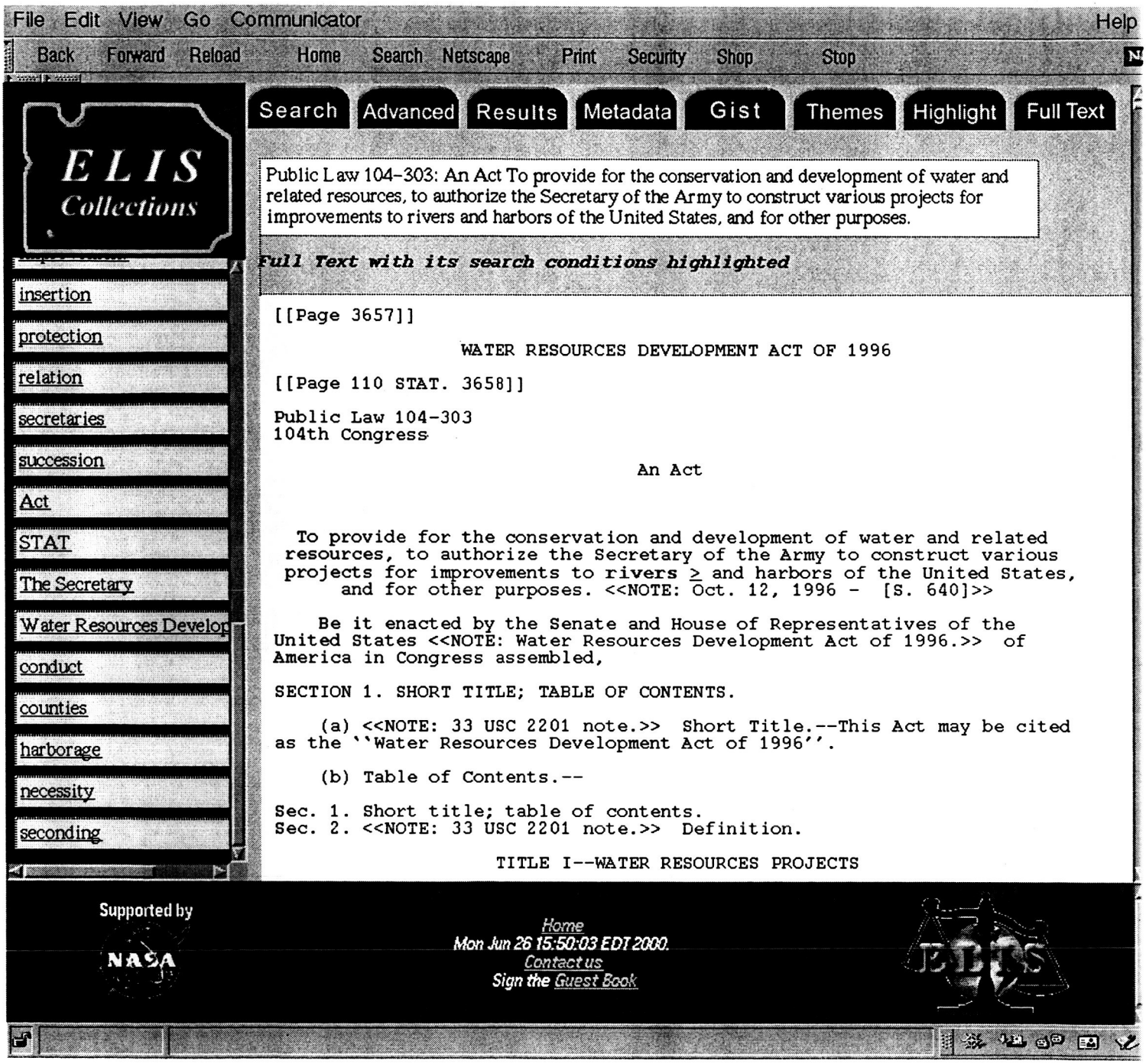

Figure 9: User requests to see the 1st hit with the terms maching his/her query highlighted and hyperlinked with each other.

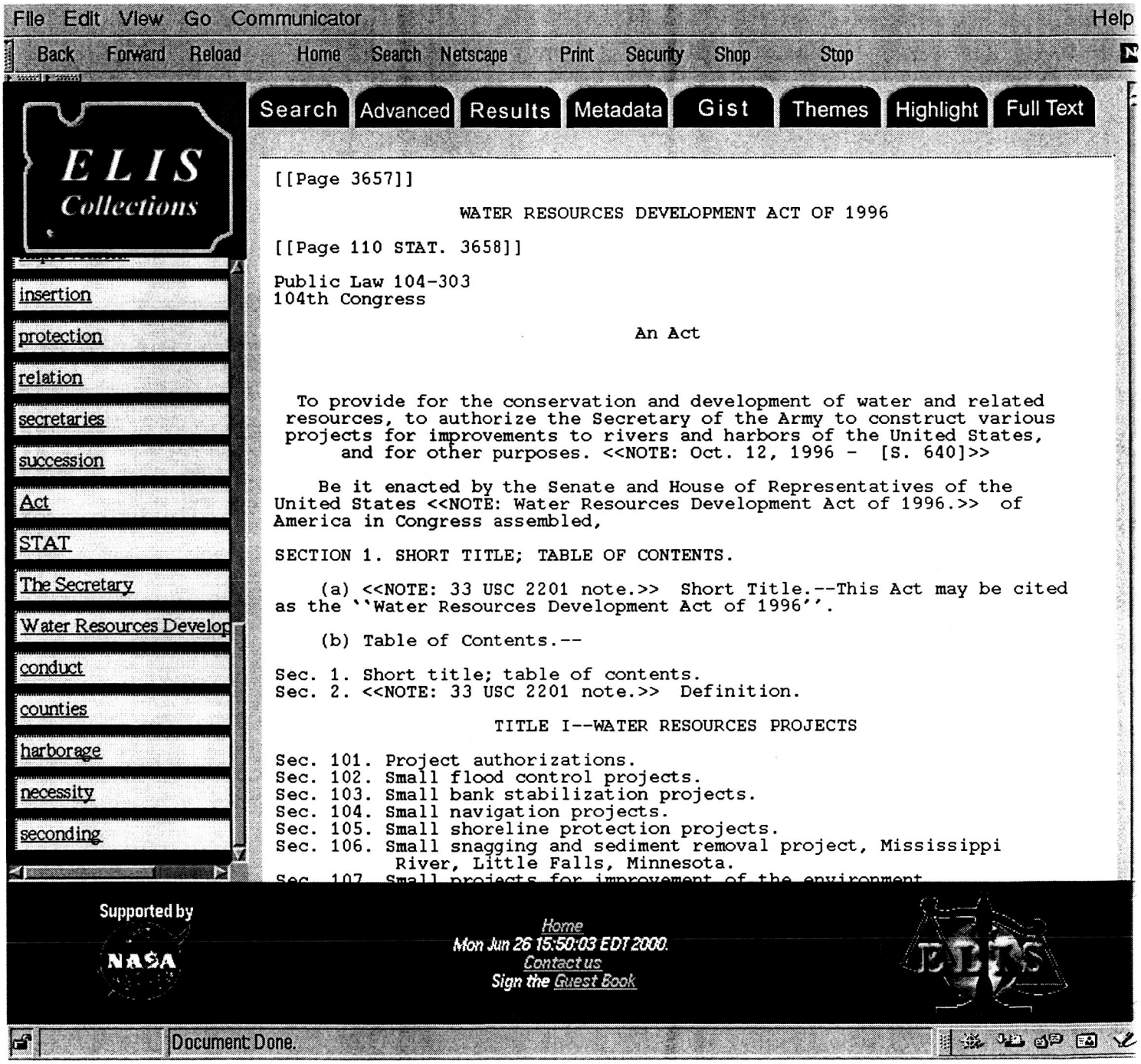

Figure 10: User requests to see the full-text of the 1st hit.

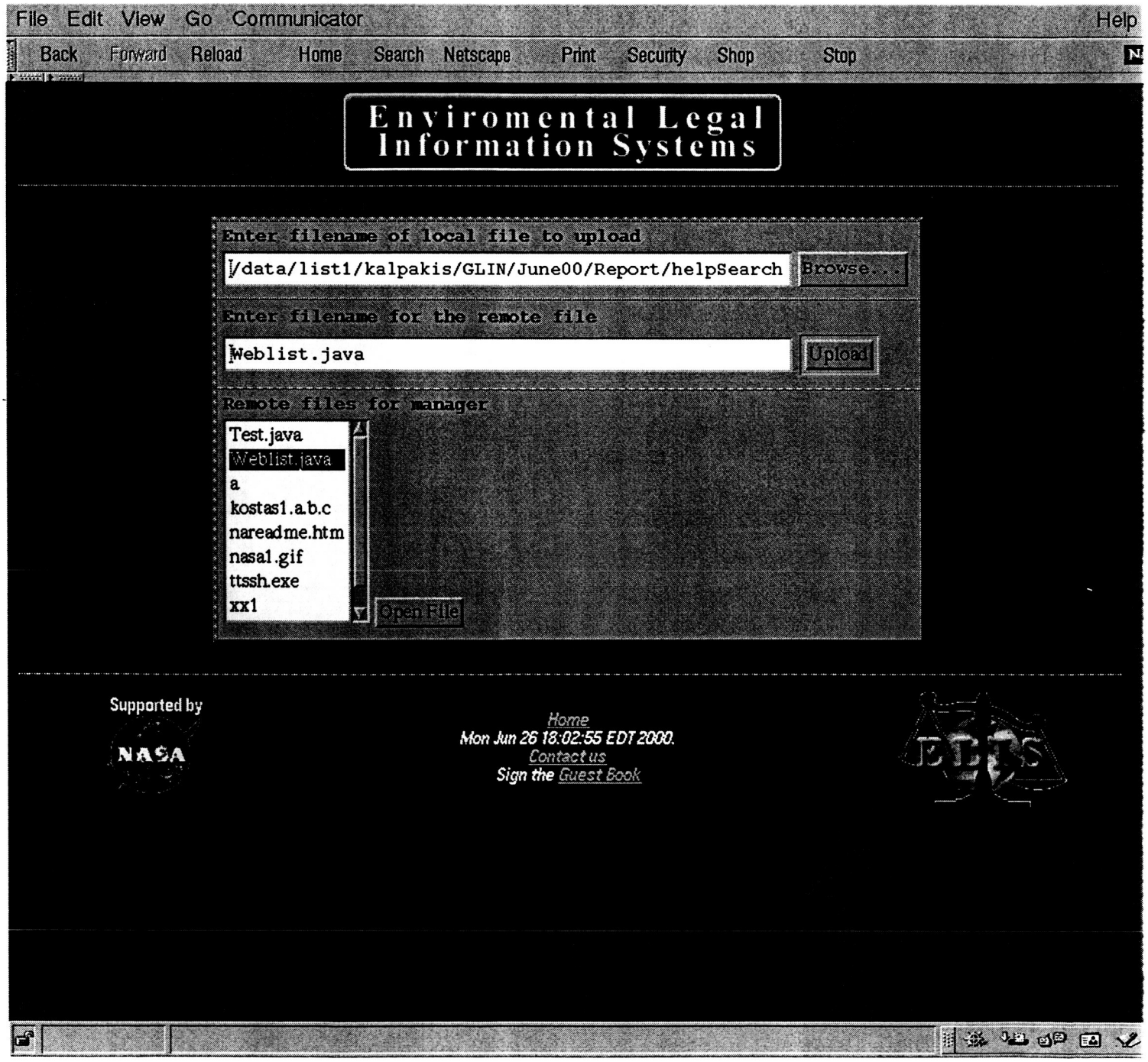

Figure 11: User uploads a local file to the remote GLIN server.

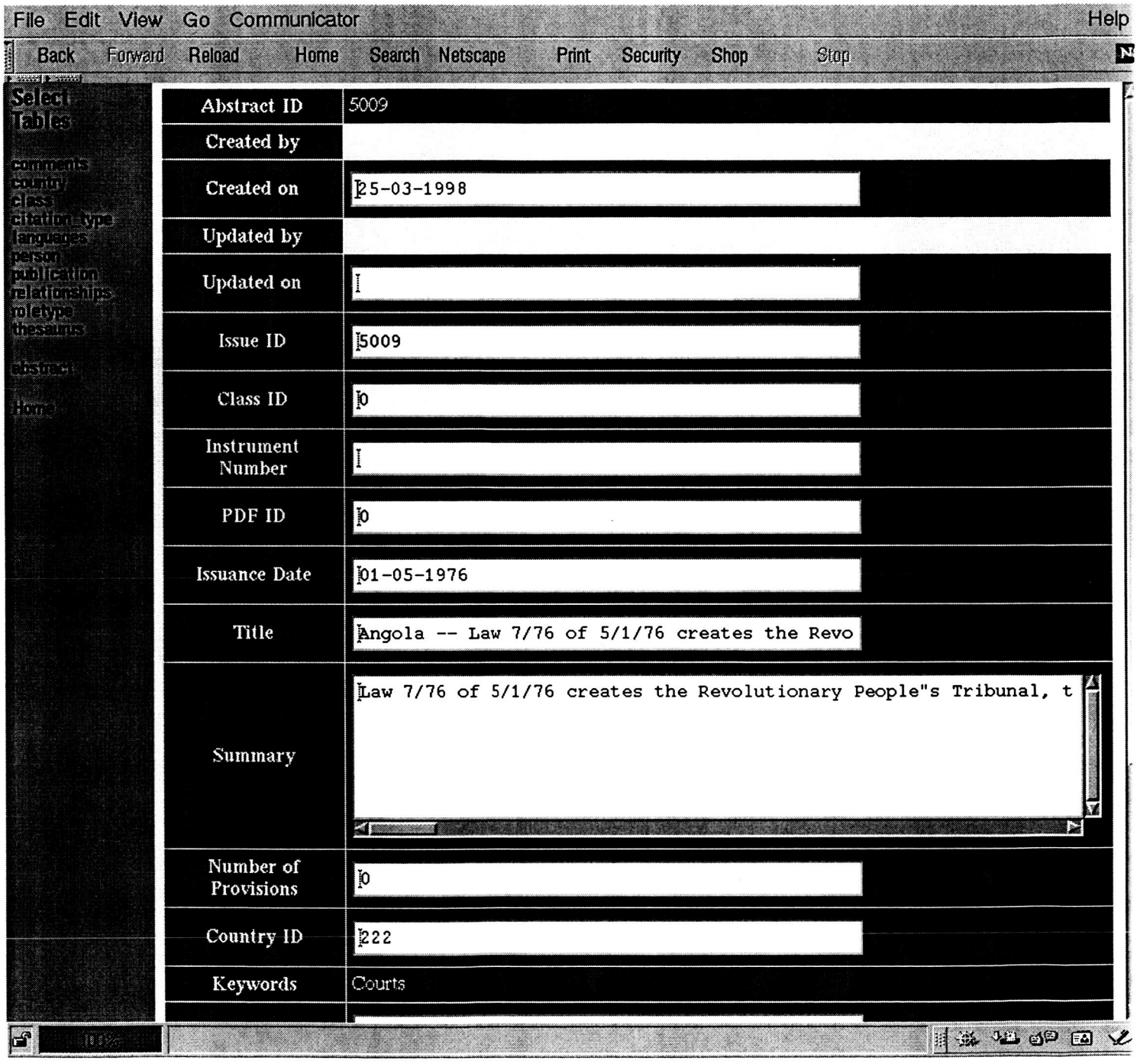

Figure 12: User edits a record for a GLIN summary document.

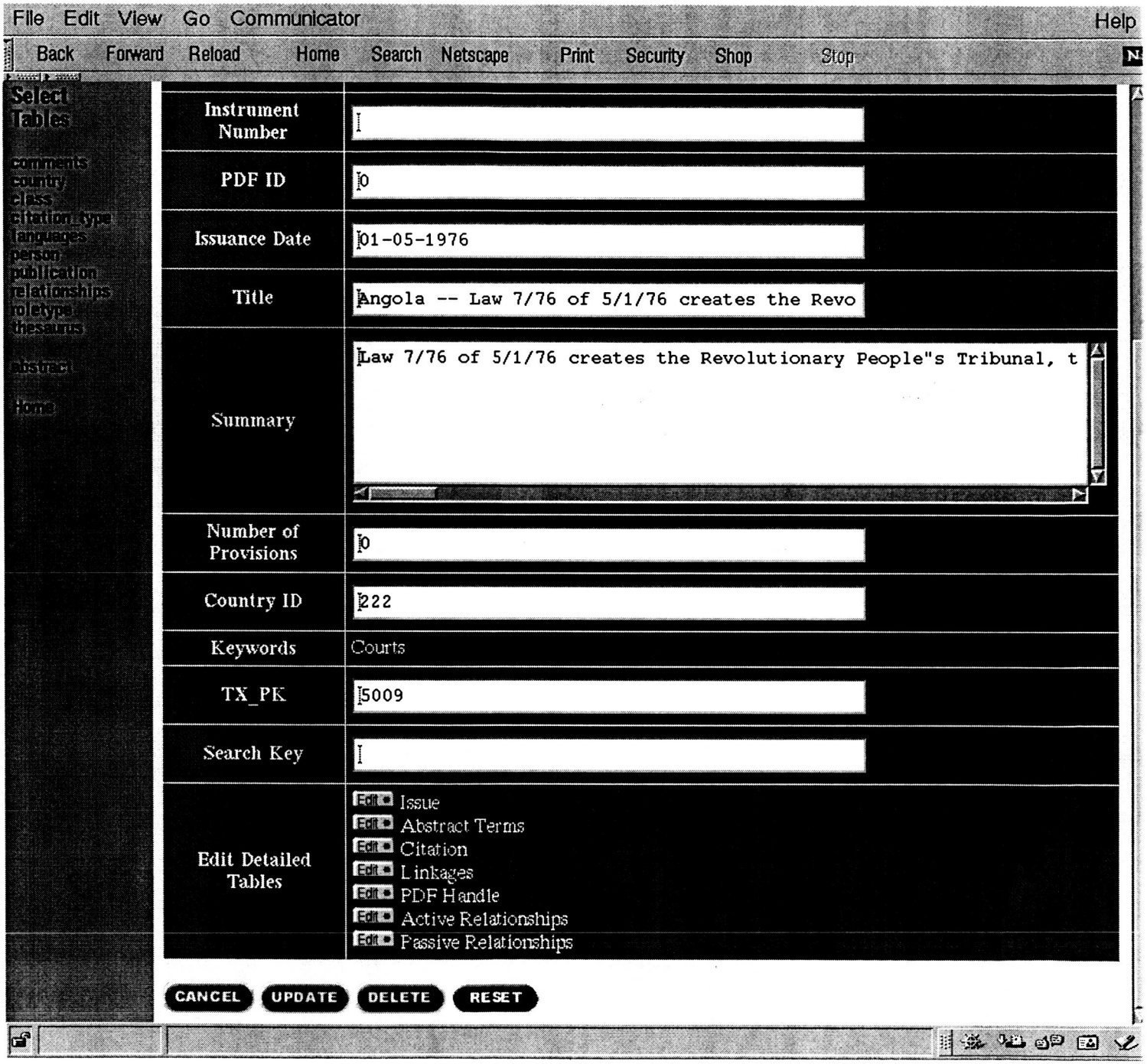

Figure 13: User edits a record for a GLIN summary document.

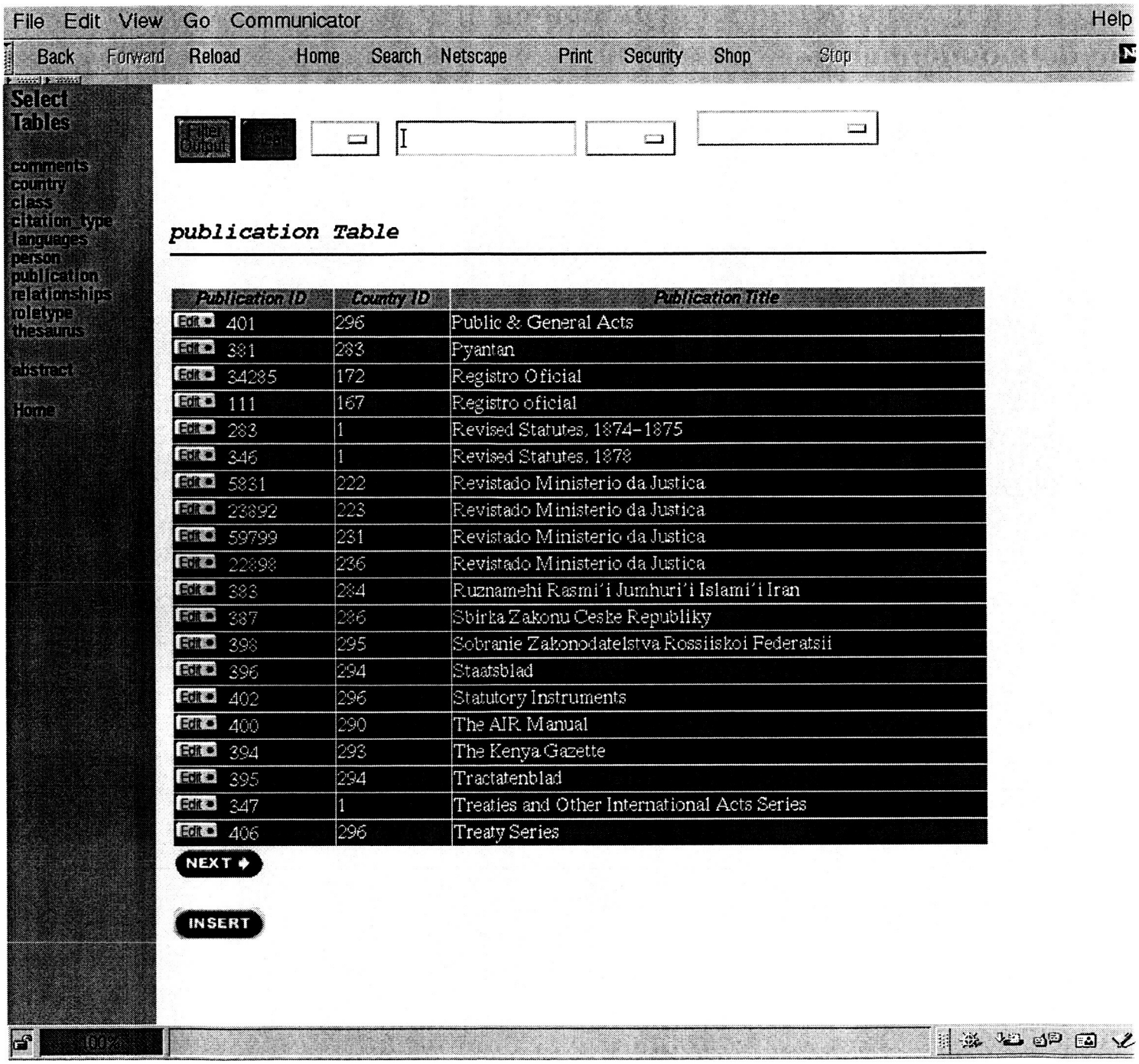

Figure 14: User edits a record of the publications table.