Research on the Model Construction and Development of Computer Information Acquisition System

Deng Rui
Xi’An International Studies University, Xi’an, China

Abstract. Nowadays, the number of crimes committed by computers is increasing, so computer forensics technology has become the focus of people's research and attention. This paper designs the module structure of the computer information acquisition system, and focuses on the realization of the key technologies of the computer information acquisition system. It discusses the application of data mining on the web, and focuses on the analysis of web text mining technology. This paper proposes a method to acquire UNIX kernel information at the user level, and introduces a utility program to acquire index node table in UNIX system. At the same time, a prototype of a professional news information acquisition system is proposed. Taking the prototype of the system as the background, the application of information acquisition on the web is realized by using relevant technologies of web information search and web text mining.

1. Introduction
With the continuous development of social informatization, the number of computer crimes is also increasing. In recent years, the development speed of illegal and criminal activities carried out by using the network is faster and faster [1]. Reliable production equipment is not only the most basic guarantee for the production process quality of manufacturing enterprises, but also an important fixed asset of enterprises. Its management directly affects the economic benefits and market competitiveness of enterprises [2]. In the process of studying and working at present and in the future, the application of computer technology is an inevitable development trend. The existing technology is difficult to meet the needs of the new situation, so it has become a very important way to use the power of law to deal with cyber crimes. At the same time, due to the long-term use of foreign software by domestic forensics agencies, this has also affected the security of state secrets to some extent [3]. This paper mainly studies the detailed design and implementation of computer information acquisition system on the basis of mastering the knowledge of operating system and file structure.

2. Overall System Design
The system studied in this topic is totally divided into two parts, namely, background server and U disk client. The background server mainly has two functions, namely, evidence customization and evidence analysis, and the customization of evidence customization includes two aspects: types and keywords. In the field of the use of computer information systems, information security is not 100% guaranteed, bringing great hidden dangers to people's information security [4]. The system mainly uses the U-disc as information acquisition equipment, and the specific functions are described as follows: the automatic playing function based on the U-disc realizes the selective acquisition of information; Secure login via ikey; Effective encryption of U-disk files; To achieve a variety of types of search. The automatic playing function of the U-disc is utilized to realize that the U-disc can
selectively obtain evidences [5]. The U-disk can directly store the information related to the client
terminal to the U-disk in a short time without being detected by other users. After the administrator
inserts the U disk into the background server, he can automatically import the relevant information in
the U disk into the server. The block diagram of the computer information acquisition system is shown in
**Figure 1**. Ikey-based secure login. Backup to disk. Full encryption. Digital signature. Search for file
names, keywords, WEB logs, etc. Its core is to quickly construct a prototype system that can work
according to users' requirements, and then modify it continuously according to users' requirements.

![Module Diagram Of Computer Information Acquisition System](image)

**Figure 1 Module Diagram Of Computer Information Acquisition System**

The standard of computer information acquisition system is described by OWL type documents, so
the essence of information acquisition is the retrieval of OWL documents. The C/S architecture can
not only include Server of various sizes and processing capabilities required by the current and future
business development of the system and their corresponding OS and DMBS, but also include various
existing network systems, host systems and file service systems of the system. Fast, according to a
predetermined task, the U-disk can quickly obtain digital evidence from the suspect computer [6].
Invisibility, the use of this system can be used to obtain evidence without the suspect being aware of it.
The Web is a highly dynamic information source [7]. The Web is not only growing at an extremely
fast speed, but also its information is constantly changing. As the soul of the Web, the smooth flow of
information is the key to activate Web applications and promote their development. It contains a large
amount of core data when the system is running. If we can get the information maintained by the
kernel in time, it will be very helpful for us to know all kinds of activity states of the whole system and
to master the running situation of the system. However, OWL-type description documents cannot be
directly used in the system and must be developed before they can be applied.

### 3. Search for Common Documents

#### 3.1 Lucene Introduction

As a method of searching all texts in a document, full-text search can scan and index each word in an
article using a computer program, and indicate the position and number of occurrences of the word in
the article [8]. Record and summarize the operation status and production data of the equipment,
conduct comprehensive statistical analysis and report output on the indicators such as equipment
operation rate and completeness rate, and compare the equipment life cycle and maintenance interval.
The method consists of a series of development stages, each of which produces a corresponding
software product. Each product provides some required operating functions, which in turn may
produce new software requirements [9]. Once users can search according to the index, they can use the
search process to search. As a software system established according to the full-text retrieval theory,
the full-text retrieval system has different functions such as indexing and adding indexes. With the
help of the document manager, the import of the ontology is realized, and the related class information and attribute information in the ontology are obtained according to the method, and the acquisition of implicit information is realized through the reasoning of the ontology. The aim is to provide developers with a simple toolkit to achieve better full-text retrieval, or to use it as the basis for a full-text retrieval engine. Realize the automatic generation of equipment maintenance plans and the intelligent management of equipment state maintenance, and standardize the corresponding treatment specifications and measures for equipment early warning, faults, accidents, etc.

3.2 Search for Text inside Pdf Files
Lucene can only process text directly, but there are still other file formats such as PDF in the computer. During the retrieval process of these files, they need to be converted into a unified format. The method regards all the development processes as consisting of a thousand spiral and repeated processes, and each process can use any one of the above methods, including various stages of software development [10]. In the process of transmitting data information, there will always be some electromagnetic waves. If the electromagnetic waves are interfered by other electromagnetic waves in some range, it will affect the data transmission process. The computer information acquisition system analyzes ontology into triplets, mainly involving three aspects: subject, predicate and object. For the above-mentioned codes, an OdModel is mainly generated through Mod-elbt Factory, and then the ontology module is transferred to the corresponding memory through relevant files. PDF format files enclose text, color, etc. in one file. If you want to extract text information, you need to analyze its file format. PDFBox, Xpdf, etc. are common parsing tools. The different values of the mode setting register “RCVALL” in the AR-CNET controller determine the information flow range intercepted and sent to the controller buffer. It emphasizes that the real world is directly mapped into software through the concept of objects, thus realizing efficient and high-quality software design.

3.3 Word Text Extraction
As a compound document, WORD document is a file system structure defined based on object connection embedding technology. In addition to format information, WORD document contains a lot of additional information. According to the location of the information code determined by the information collection dictionary in the data domain and its characteristic combination as the core, a system identification module is compiled. After the data domain is identified, the equipment information data is stored in the field information database. Database applications based on C/S generally span two hardware platforms. Therefore, we should first consider distributing the programs and data of the database application system on different computers. As an open source library of Apache Software Foundation, Apache POI can provide the Java program with the reading and writing function of files in Microsoft Office format. The POI in Apache can be used to extract WORD and Excel documents. The user terminal can generate relevant interception conditions according to the form of keywords or ontology corresponding trees. Secondly, through ontology analysis, the positioning system captures the category information corresponding to the conditions. When dividing a database, we should pay close attention to the interaction between entities. Generally, as much locality as possible should be used. Most systems do not always refer to locality and need to strike a balance between centralized and distributed access. The controller is set to “Receive All” mixed mode, and the information acquisition unit ignores the destination address of the datagram, directly listens to all datagrams transmitted on the network, and decodes and analyzes them.

4. The Development of Utility Inodestat
In UNIX system, the logical expression of a file is completely different from its physical expression. The logical expression of a file refers to the file composed of characters that users see after typing cat commands, while the physical expression of a file refers to the actual storage of the file on disk. Due to the interconnection between documents, useful information other than document contents can be provided. Using this information, you can sort the pages and find important pages. With POI, POI can
directly access the file format of WORD. Lucene and POI are subprojects of Apache, so they can connect seamlessly. There are two types of object structures: classification structure and assembly structure. The classification structure can abstract common features from objects to form classes in the problem space, and can be extended to instances. It illustrates the generality and specificity among objects. It can be reconfigured for each user at regular intervals. When a user accesses a Web document, he usually accesses a plurality of Web documents continuously along the super connection, and these web documents actually indicate the current situation of the user. Therefore, when a user accesses the file, the UNIX file system needs to spell the blocks storing the file together in order to give the logical expression of the file required by the user. The system can expand the semantic of keywords to generate upper keywords and lower keywords, thus expanding the scope of information acquisition.

For this reason, a table is set up in UNIX system. The file system converts the physical expression of the file into logical expression according to this table. This table is called index node or I node. Each index node is a 64-byte table that stores information related to files. The specific implementation method is to obtain the first address in the URL list, obtain the corresponding Web document for preprocessing, and find out the hyperlink to other Web documents in the document. Using the method of COM-Java, Microsoft's WORD program can exist in the form of COM. If COM components can be called, they can be used to obtain text information. In the analysis process of determining the connection between objects, object-oriented analysis only establishes a communication mechanism between objects to form message transmission between objects, thus reducing the connection between objects and thus reducing the coupling between system modules. Such as file length, file owner, file permissions, file type, etc. In addition, another important item is the disk address table, which gives the real physical address of the file. Its task is to understand the objects involved in the problem in the problem domain and the relationships between the objects to build the problem model. Considering that it takes a long time to construct an association database, which will affect the current speed of users accessing Web pages, the user's access trajectory is first saved and then constructed once at a certain moment.

![Figure 2 Reverse Maximum Matching Algorithm Flow](image)

Every file in UNIX system has a unique index node. The index node is stored on disk in static form, which is called disk index node. When the UNIX kernel wants to perform various operations on a file, it reads the corresponding index node information into memory. Robot of this system also expands according to the synonyms and words of keywords in the knowledge base, automatically searches web documents from the designated URL by using the depth-breadth first combination search strategy, and returns the searched results to the decision-making module for processing. In the analysis phase, OMT emphasizes to start with the problem statement and establish corresponding models for the objects in the system, the relationships among the objects, the event flow and the functions respectively, namely
the object model, the dynamic model and the function model. These models comprehensively reflect the requirements of the system from different angles. The ontology model is acquired through information query language. Information acquisition is to match the triplets as query conditions with the triplets of the ontology model, and the query result is the result of matching the condition triplets with the triplets in the ontology model. According to statistics, the error rate of forward maximum matching is 1/169, and the error rate of reverse maximum matching is 1/245, therefore, the system adopts reverse maximum matching algorithm for Chinese word segmentation RMM algorithm flow chart is shown in Figure 2. Each item in the file table has a pointer to a specific item in the index node table and a file read/write pointer. The real purpose of setting the file table is to save the file read/write pointer. The data area of each active process contains a pointer that points to the file table. The pointer of the file table points to the index node, and the pointer of the index node points to the file.

5. Conclusion
In a word, with the development of network and information technology, the application of computer information acquisition system will be more and more extensive. Through the development of computer information acquisition system, people can acquire real-time information of network terminals. This system combines the search engine technology and web data mining technology on the internet, so that users can get the information that is closest to their desired target and is really interested in from the search results. Obtaining index node table this is only an example of the method of obtaining UNIX system kernel information described in this paper. Using this method, we can also obtain a large number of kernel data structures such as active process table and file system installation table and various state information during system operation. We can master the terminal state and relevant contents of the information, thus better locating the contents of interest to customers and avoiding the occurrence of network security risks to a certain extent.

Acknowledgement
Supported by the educational reform fund of Xi’an International Studies University (no: 18BZ11)

References
[1] Xiao K, Du Z, Yang L. (2017). An Embedded Wireless Sensor System for Multi-Service Agricultural Information Acquisition[J]. Sensor Letters, 15, no. 11, pp. 907-914.
[2] Erdenebat M U, Kim B J, Piao Y L, et al. (2017). Three-dimensional image acquisition and reconstruction system on a mobile device based on computer-generated integral imaging[J]. Applied Optics, 56, no. 28, pp. 7796.
[3] Langer M, Cornelius J. König, Fitili A. (2017). Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection[J]. Computers in Human Behavior, 81, pp. 19-30.
[4] Osipov V Y, Vorobiev V I, Levonevskiy D K. (2017). Problems of Protection against False Information in Computer Networks[J]. Tr Spiiran, 4, no. 53, pp. 97–117.
[5] Shen G. (2018). Computer and Information Technology, Firm Growth, and Industrial Restructuring: Evidence from Manufacturing in the People's Republic of China[J]. Social Science Electronic Publishing, 35, no. 1, pp. 133-154.
[6] Lee R. (2018). Computer and Information Science[J]. Studies in Computational Intelligence, 719, no. 5, pp. 74.
[7] Lorence D, Jones A. (2017). A wholistic model for teaching computer and information science: learning on a WhIM[J]. International Journal of Innovation & Learning, 4, no. 4, pp. 376-390.
[8] Fallavollita P, Kobayashi E, Speidel S. (2018). IPCAI 2018 Special Issue: Information Processing for Computer-Assisted Interventions, 9th International Conference 2018-Part I[J]. International Journal of Computer Assisted Radiology and Surgery, 13, no. 5, pp. 607-610.
[9] Li T. (2017). Analysis of Computer Network Information Based on “Big Data”[J]. IOP Conference
[10] Gangeh M J, Member S, IEEE, et al. (2018). Computer-Aided Theragnosis Based on Tumor Volumetric Information in Breast Cancer. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 65, no. 8, pp. 1359-1369.