Design and Application of Intelligent Forensic Medicine Comprehensive Application Platform Under the Background of Big Data

Lugui Hao(✉)

Sichuan Judicial and Police Officers Professional College, Deyang 618000, Sichuan, China
957670052@qq.com

Abstract. With the vigorous development of digital information technology, big data, as a new means to adapt to the development of the times, has become an important force to promote the digital and intelligent transformation of forensic identification work mode. In this regard, this paper takes the practical application of forensic identification as the research object, and integrates the research results in the fields of big data technology, network information technology and computer application technology to build a comprehensive application platform for intelligent forensic medicine. The platform takes Hadoop cluster as the core to build a data storage, analysis and processing server, which highly collects medical data, traffic monitoring video data and other graphic information data. With the help of Lucene algorithm tools and BoVW scheme, the design and development of search engine are completed to meet the needs of users. In addition, the platform will use Javaweb technology to complete the design and development of the front-end of big data visualization interaction, form a standard Web application, and provide users with convenient and efficient data services.

Keywords: big data technology · forensic identification · Hadoop · search engine · computer application

1 Introduction

As the main executor of national judicial expertise, forensic doctor is an indispensable part of modern public security criminal investigation system. Its duty is to comprehensively use medical technology, criminal science technology and forensic expertise technology to provide necessary help for restoring the process of the crime, sorting out clues and evidence, and making a detection plan [1]. Common forensic identification includes pathological identification, clinical identification, mental identification, physical evidence identification and poison identification, and it is open to the whole society, so that the scope of forensic work is becoming wider and wider [2]. At the same time, the work of forensic identification is demanding, the technical and professional requirements are complex, especially the current forms of crime are increasingly diversified, and the difficulty of forensic identification is also increasing. In addition, the work mode of forensic identification involves many links, which is not only time-consuming and

P. Qi and Z. Chen (Eds.): ICBDIE 2023, AISR 178, pp. 110–114, 2024.
https://doi.org/10.2991/978-94-6463-238-5_15

labor-intensive, but also has insurmountable subjective factors [3]. In view of this, this paper believes that the construction of intelligent forensic comprehensive application platform under the big data environment is of positive significance to promoting the informatization construction of forensic identification, improving the detection of cases, the quality of identification and the efficiency of forensic work.

2 Development Process

First of all, the construction of Hadoop cluster architecture needs the support of hardware settings and software programs. In terms of hardware equipment, according to the functional requirements of the platform, Hadoop cluster includes five nodes, named as Master1, Slave1, Slave2, Slave3 and Slave4 respectively. Each node needs two quad-core hexadecimal CPUs or one eight nuclear hexadecimal CPU, with a running frequency of 2.5 GHz and a memory of 512G. The hard disk needs 12–24 batch disks with a capacity of 1-4TB to meet the storage requirements of a large amount of data [4]. As for the software program, Linux is selected as the bottom operating system, CentOS 7.6 (x86_64) as the version, jdk-8u291-linux-x64 as the version of JDK, and Hadoop as the version of 3.3.1 [5].

Secondly, the rapid retrieval and call of massive data information needs the help of a retrieval engine. In the face of structured data, the platform will support Lucene technical framework to complete index construction and retrieval algorithm model. The following is the key code of Lucene technical framework to construct the index. For monitoring video, images and other unstructured data information, the platform provides a visual bag of words (BoVW) model to complete the retrieval function. In the BoVW model, the SIFT feature of each image is transformed into the word frequency vector representation of the image, and the TF-IDF algorithm is used to weight the word frequency vector, which simplifies the image representation and provides convenience for subsequent retrieval [6].

```
Directory directory = new SimpleFSDirectory("/Users/chinatsui/tmp/index");
Index Writer indexWriter = new IndexWriter(directory, new
StandardAnalyzer());
index Writer.setUseCompoundFile(false);
Document doc = new Document();
doc. add(new Field("title", "This is the title text", Field.Store.YES,
Field. IndexNOT_ ANALYZED));
doc. add(new Field("content", "This is the content text", Field. Store.YES,
Field. Index. ANALYZED));
index Writer. addDocument(doc);
```

Finally, for the construction of Web Server, Java is selected as the basic development environment, MyEclipse V 2022 as the integration tool, Tomcat 8.0 as the Web server and MySQL 5.7 as the database server. Through the introduction of the above key technical theories, the overall environment of system development, the configuration of related

software and tools are determined, and the technical feasibility of the overall project of intelligent forensic comprehensive application platform is also clarified.

3 Functional Implementation

3.1 Data Information Upload

After logging in to the system, users can upload the data of various types of inspection and identification, and classify them according to pathological identification, clinical identification, mental identification, material evidence identification and poison identification. The platform supports Input upload control, drag and drop file upload, paste file upload and other ways [7].

3.2 Comprehensive Comparative Appraisal

Under this function module, forensic experts can quickly retrieve the data information or video images needed for inspection and identification with the help of the platform's high-speed retrieval engine, which provides convenience for the determination of time of death, individual identification and the nature of death in forensic identification. For example, in the examination of determining the time of death, the traditional working mode mostly relies on postmortem phenomena, autopsy and the state of bones to infer the time of death, especially in the early stage after death, and there are often some errors [8]. In this regard, forensic experts can directly input the appearance image of the deceased in the platform, and the platform will automatically extract the SIFT features from the input image by using the BoVW model, and complete the word frequency vector transformation by using the local features of the image according to the visual dictionary. Then the TF-IDF formula is used to calculate the weight value, and the parallel retrieval of images is completed, so that the relevant traffic monitoring video data can be obtained quickly. Formula 1 is the formula for calculating TF-IDF, where $n_{i,j}$ stands for the number of times a word appears in a document, $\sum_k n_{k,j}$ is the sum of the number of times all words appear in a document, $|D|$ is the total number of documents, and $\left|j \in d_j\right|$ is the total number of documents containing this word [9].

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}},\ IDF_i = \log\frac{|D|}{1 + \left|j \in d_j\right|},\ TF - IDF_{ij} = TF_{i,j} \times IDF_i \tag{1}$$

The inverted parallel retrieval of images needs to rely on MapReduce under Hadoop. The Map function calculates the weighted BoVW vector of the query image, and finds the candidate video image set according to the inverted index file. The Reduce function calculates the similarity between the candidate video image set and the query image, and outputs the similarity values in descending order as the final retrieval results. The retrieval results of BoVW model images are shown in Table 1.

In addition, in the process of clinical examination, forensic appraisers can input keywords of clinical examination information in the platform to retrieve the related medical information of the injured, which reduces the risk of distortion of forensic examination results. During keyword retrieval, Lucene framework will use several APIs such as

Table 1. Image retrieval results of BoVW model

Input image	Retrieval threshold	Total similarity	Retrieve similar quantity	Accuracy rate	Coverage rate
P91	20	30	12	60.00%	0.400
P337	20	55	19	95.00%	0.350
P1152	20	40	11	55.00%	0.275

Table 2. Text retrieval functional test results

No.	Retrieval time	Thread	Tab	Connect time (ms)	Delay time (ms)
1	12:57:03	1–1	Network request	11	83
2	13:01:22	1–2	Network request	20	107
3	14:33:25	1–3	Network request	13	90

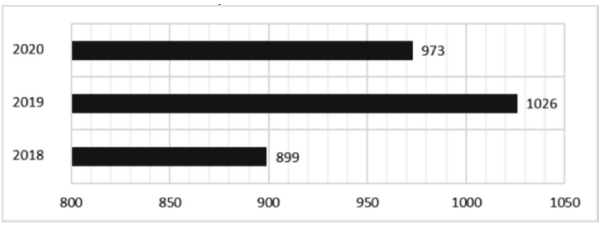

Fig. 1. Summary of forensic examination and identification of rape cases

IndexSearcher, Query, QueryParser and Filter to perform keyword query operation on the index database, as shown in Table 2, which shows the test results of text retrieval function [10]. The results show that Lucene framework can support high concurrent retrieval, and the average response time is below 100ms, and its performance meets the design specifications.

3.3 Data Statistical analysis

Under this function module, forensic experts can make statistical analysis of all cases in a certain period of time. Figure 1 shows the forensic examination and identification of rape criminal cases in recent three years.

4 Conclusion

In order to promote the reform of forensic examination and identification work mode, this paper aims at many shortcomings in the traditional mode, and builds a comprehensive application platform of intelligent forensic medicine with the help of the practical

characteristics of big data technology, network information technology and computer application technology. The platform can not only support the digital upgrade of massive forensic information, strengthen the sharing and flow of data information, but also collect medical data, traffic monitoring video data and other graphic information data, which provides convenience for forensic examination and identification and promotes the digital and intelligent transformation of forensic identification.

References

1. Wang Hailong, Lin Xi'nai. The Decipher of "Death Code" [J]. Road Traffic Management. 2022.01.
2. Fang Yating, Lan Qiong, et al. New Opportunities and Challenges Faced by Forensic Science in the Era of Artificial Intelligence Technology [J]. Journal of Forensic Medicine. 2020.02.
3. Jiao Yanning. Analysis of the Current Situation of Forensic Medicine in Grass-roots Public Security Organs and Its Countermeasures [D]. Zhengzhou University. 2021.09.
4. Li Li. Attempt and Thinking of Deploying Hadoop Cluster under Hyper-converged Architecture [J]. Modern Instruments & Medical Treatment. 2021.01.
5. Shi Fangxia, Gao Yi. Application Analysis of Hadoop Big Data Technology [J]. Modern Electronics Technique. 2021.09.
6. Wang Qian, Chen Yiming, et al. Vehicle Recognition Algorithm Based on Visual Bag Model in Complex Environment [J]. Journal of Computer Applications. 2018.05.
7. Ruan Xiaolong, Li Pengnan. Efficient Uploading Method of Large Files Based on Web [J]. Electronic Technique. 2020.03.
8. Xia Zhiyuan, Cong Bin. Research Progress of Early Death Time [J]. Chinese Journal of Forensic Medicine. 2018.04.
9. Chen Ruiwen. Research on Image Retrieval Method Based on Improved BoVW Model [J]. Journal of Chongqing University of Science and Technology (Natural Sciences Edition). 2015.05.
10. Jia Xiaoxia. Design of Dynamic Retrieval System of Network Resource Index Information Based on Lucene [J]. Microcomputer Applications. 2021.03.