International Journal of Future Generation Communication and Networking
Vol. 9, No. 3 (2016), pp. 1-14

Design & Research of Legal Affairs Information Service Platform Based on UIMA and Semantics

Wang Chen, Zhang Jia and Qin Lele*

Hebei University of Science & Technology, Shijiazhuang, China
Mr_qin@163.com

Abstract

The law is the most powerful weapon to safeguard national stability and ensure flourishing of all causes as well as an instrument to protect the rights and interests of the masses, therefore, the just and accurate use of legal provisions is of crucial importance. With the increase of informatization level of the legal industry itself, more and more services in legal affairs are provided in information-based forms, and there is a large number of unstructured information in such business data. It's a problem the legal industry needs to tackle in information management to rapidly acquire valuable content from the massive unstructured data and make use of such content. Based on analysis of problems arising in existing laws and regulations informatization system, this paper comes up with solution of a legal affairs information service platform based on cloud computing, UIMA, semantics, big data and Chinese word segmentation. This paper also proposes the four-layer technical framework solution on the basis of the design of integration and management method of unstructured data of heterology and isomerism, analyzing and processing method of unstructured data, semantics based unstructured information retrieval method and construction and maintenance method of ontology library. It also provides detailed introduction to the realization of the combination of big data and cloud computing and its application in this information platform by virtue of designing UFS - a distributed file system, MapReduce - a batch processing technology and BigTable - a distributed database. Data acquisition and expanded data analysis can be conducted by making use of the expandable UIMA framework and the sequential indexing of data content and analysis results can be materialized by means of applying Lucene indexing technology. With regard to information retrieval, the concept of ontology is introduced on the basis of traditional search model and a new search model based on domain ontology is proposed. IKAnalyzer 3.x is proposed to facilitate Chinese word segmentation. By taking advantage of such information service platform, legal affairs enterprises can effectively integrate structured and unstructured information resources and implement the storage, analysis, retrieval and decision-making applications of business data content.

Keywords*: Legal affairs, UIMA, Big data, Chinese word segmentation*

1. Introduction

Generally, the law is a code of conduct under social recognition and state affirmation with the norms formulated by legislative bodies and implementation safeguarded by coercive force of the state (mainly judicial offices). It stipulates the rights and obligations of the parties involved with constraining force to all members of society. The law is the most powerful weapon to maintain state stability and prosperity, a tool to safeguard the rights and interests of the masses, and an organ of dictatorship through which the governors rule the ruled. The law is a series of rules which are usually implemented through a set of institutions. The implementation of rule by law is a

ISSN: 2233-7857 IJFGCN

significant symbol of social civilization as well as an important guarantee for long-term state stability. Therefore, it's crucial to utilize legal provisions in a just and accurate way. In legal research, teaching and legal practice, the retrieval of laws and regulations and relevant cases is an important groundwork. However, there is a so large number of laws and regulations and regulatory documents, and relevant legal cases, which are mainly stored in unstructured ways such as audios and videos, are different from each other, therefore, it's really hard to find the required information accurately. But, with the development of modern information technology, "to look for a needle in the ocean" is no longer a myth. By making advantage of data computing methods emphasized in cloud computing, massive information storage, analyzing and processing ways emphasized in big data technology and the unstructured information management system framework in UIMA, a legal affairs information service platform can be built to complete tasks such as massive storage of unstructured data including texts, audios and videos, as well as to conduct retrieval and decision-making support by means of word segmentation, legal reasoning technology and bit notation, reducing the time-consuming and toilsome searching burden into a simple click of the mouse.

2. Problems of Existing Laws and Regulations Retrieval System Introduction

Since the first laws and regulations retrieval system came into being in 1985 in China, a big bunch of retrieval systems are published one after another, including both comprehensive retrieval systems such as Law & Regulations of China by State Information Center and specialized retrieval system such as Sanitation Regulations, Standards and Test Methods by Department of Law Enforcement and Supervision of the Ministry of Health; information can be retrieved from a CD such as FWT Library of Laws and Regulations by Beijing Fawutong IT Co., Ltd. or via the internet such as Chinese Civil and Commercial Law Retrieval System by Chinese Civil and Commercial Law website, or by applying both ways such as Chinese Law Retrieval System by ChinaLawInfo. Co., Ltd. It has become challenging to find the laws and regulations we required quickly from the many good and bad retrieval systems. The disadvantages are as follows:

(1) Incomplete Data

Complete data are supposed to include all laws and regulations issued by the state, in effect or out of work, corresponding cases and analysis. Though laws and regulations in effect are essential, those out of work are important for legal research and corresponding cases and analysis are interpretations of the provisions. Therefore, the absence of any aspect leads to data incompletion. It is a pity that no one among the numerous existing laws and regulations retrieval systems deserves to be called complete so far.

(2) High Data Repetitive Rate

Repeated data bring uncertainty and conceptual confusion to users and thus harm the authority and credibility of the retrieval system. For example, if we retrieve "copyright law" in Collection of Chinese Laws and Regulations by China Jilin Beicheng Company, we will find three entries of copyright laws with the same content but different indexes.

(3) Non-Standard Classification

This is the common weakness of all laws and regulations retrieval systems. Each retrieval system classifies the laws and regulations included in accordance with its own understanding of legal categories rather than a uniform standard, leading to different numbers of classifications. Such condition leads to deviation or dissatisfaction of users during the retrieval process. Besides the lack of uniform standard for classification, other evidences include non-standard category names and non-standard classification.

(4) Insufficient Relevant Cases for Reference

As is known to all, different judges and lawyers have different understandings and sentencing criteria of laws, resulting in a great deal of reference materials, especially relevant video, image and audio documentations. However, there are still few legal affairs inquiry platforms integrating both legal provisions and relevant cases at present.

(5) System Undercapacity of Shared Data

The construction of a legal affairs service platform integrating legal provisions, manner of writing and typical case analysis (audio, video, image and text documentations included) requires big volume data storage, high-speed operation and large-scale sharing. However, there is evident deficiency of shared data capacity and materials.

(6) Young Age Of Unstructured Information Management

The coming of an era of high speed, high sharing level and big data causes the rapid increase of unstructured information resources. However, it's very hard to find interesting information and acquire underlying valuable knowledge from the massive, dynamic and isomerous unstructured information resources. Traditional legal affairs information resources mainly involve structured data such as texts; with the great development of science and technology today, various multimedia technologies have generated digitized, shared and integrated data, enhancing the significance of unstructured information resources management. At present, unstructured information management systems relating to legal affairs are still at their initial stage of development.

3. Current Research at Home and Abroad

The platform integrates cloud computing, big data, UIMA and intelligent word segmentation into legal affairs information service as they become more mature after years of application. The concept of big data is mentioned by more and more people in more and more occasions in recent years, usually connected with could computing, and the mutual application between cloud computing and big data has become a hot topic. In Big Data Using Cloud Computing published in 2014, He Qing concluded that big data generated at high speed can only be processed within reasonable time by means of cloud computing. Cloud computing is feasible to improve analyzing and understanding of big data. It's the core support technology of big data processing as well as the mainstream method of big data mining. K. Shim, F. Gao and Liu Zhihui have all wrote papers indicating the promising future of big data and its win-win combination with cloud computing. With regard to the lack of access to a large number of unstructured information in information management, IBM proposed the Unstructured Information Management Architecture (UIMA) which can analyze unstructured contents. Li Yan, Cao Meizhen, Zhang Longbao and Shao Longbin have once put the technology into application and gained good results. Many methods and algorithms are proposed in word segmentation technology, such as the Double-Character-Hash-Indexing proposed by Li Qinghu, the Chinese word segmentation method based on character-word joint decoding proposed by Song Yan and the one based on effective substring tagging proposed by Zhao Hai. The satisfactory test results of all methods show that these technologies are becoming more mature. However, they are rarely applied in legal affairs information service, indicating a vast room for research and high application value of a legal affairs information service platform based on cloud computing and UIMA.

4. System Design

With regard to the lack of access to a large number of unstructured information in information management in the legal industry, this paper comes up with the

construction of a legal affairs information service platform based on cloud computing and UIMA as a solution. Main requirements can be reduced to four aspects considering the characteristics of the legal industry:

a. Realization of Integration and Management of Unstructured Data of Heterology and Isomerism

A uniform information management platform can be built according to the heterology of unstructured data distribution and isomerism of information storage format in the legal industry.

Requirements include: (1) data resources are expandable as required when the system searches data; (2) data format is expandable as required when the system acquires data; (3) the system can work at preset time; and (4) the system conduct uniform storage of acquired data.

b. Realization of Analyzing and Processing of Unstructured Data

The system can analyze information and conduct business-related data analyzing and processing according to the characteristics of legal business.

Requirements include: (1) the system is equipped with an expandable content analysis framework; (2) the system can conduct data classification according to the content; (3) the system can work at preset time; and (4) a content analyzer changeable as required is available.

c. Realization of Semantics Based Unstructured Information Retrieval

Semantics based unstructured information retrieval can be realized in view of the characteristic of massive unstructured data in the legal industry. Traditional full-text retrieval technology enables quick search of matching materials according to keywords, though, there are defects such as excessive irrelevant information and failure in the reasoning and mining of needed information. Therefore, semantics based information retrieval plays an important role in legal affairs information management.

Requirements include: (1) realization of indexing of unstructured information; (2) realization of full-text retrieval of unstructured information; and (3) realization of semantic analysis of retrieved content.

d. Realization of Construction and Maintenance of the Ontology Library

Ontology is the formal description of knowledge and ontology library is the assembly of formal descriptions of actual knowledge. The system needs to build an ontology library to guarantee the ontology-related inquiry of keywords.

Requirements include construction and maintenance of an ontology library.

System users can be classified into four categories according to the above requirements, namely, information searchers, system administrators, ontology library administrators and business component developers.

(1) Information searchers: acquiring required data content from unstructured information resources through the inquiry function provided by the system.
(2) System administrators: responsible for user account management and access control of the whole system.
(3) Ontology library administrators: having good knowledge of legal business, constructing and maintaining ontologies within the ontology library.
(4) Business component developers: developing new content analyzing components as required and responsible for the deployment and maintenance of such components.

e. Framework Design

To meet the requirements of the four aspects, this paper proposes an unstructured information retrieval program based on UIMA and semantics.

Functions provided by the program include: (1) acquisition of unstructured information; (2) storage of unstructured information; (3) analysis of unstructured information; (4) indexing of unstructured information; and (5) semantics based unstructured information retrieval.

The program firstly integrates data sources of unstructured information of heterology and isomerism in the legal industry, and carries out uniform management of information resources by means of Content Management System (CMS). Then expandable UIMA framework is employed to conduct data acquisition and expanded data analysis of such unstructured legal affairs information resources, and Lucene indexing technology is utilized to conduct sequential indexing of data content and analysis results. With regard to information retrieval, the concept of ontology is introduced on the basis of traditional search model and a new search model based on domain ontology is proposed, and semantics based information retrieval can be realized by building a legal affairs domain ontology library based on OWL standard.

On the basis of the semantics based unstructured information retrieval program, this paper proposes the core of the legal affairs service platform - design proposal of semantics based legal affairs unstructured information retrieval system, and the platform is called LASP (Legal Affairs Service Platform). As shown in Figure 1, LASP employs a four-layer technical framework, namely from the bottom layer to the top layer:

(1) Data Persistence Layer

This layer is responsible for the persistent storage of unstructured information data. Oracle database is employed for the storage of data of heterology and isomerism as well as storage of the ontology library. Storage of file indexing is implemented through the setting of storage path.

(2) Technical Support Layer

The technical support layer is responsible for the implementation of lower-layer technologies and the encapsulation of third party API (Application Program Interface). Services in this layer are invoked to guarantee the correct execution of businesses in the business logic layer. APIs encapsulated in this layer are mainly UIMA API, Lucene API, Jena API, employee system API and CMS API.

(3) Business Logic Layer

This layer is to provide services corresponding to requests and such services are usually deployed in a logic server support by Tomcat technology. These services mainly include: information collection service, information storage service, data analysis service, data indexing service, data retrieval service, logical reasoning service and ontology library editing service.

(4) Presentation Layer

Browsers are utilized to allow users into this layer, invoke its business logics and present results. The presentation layer is deployed in Web servers supported by Tomcat technology and it interacts with the business logic layer by means of HTTP and RMI (Remote Method Invocation).

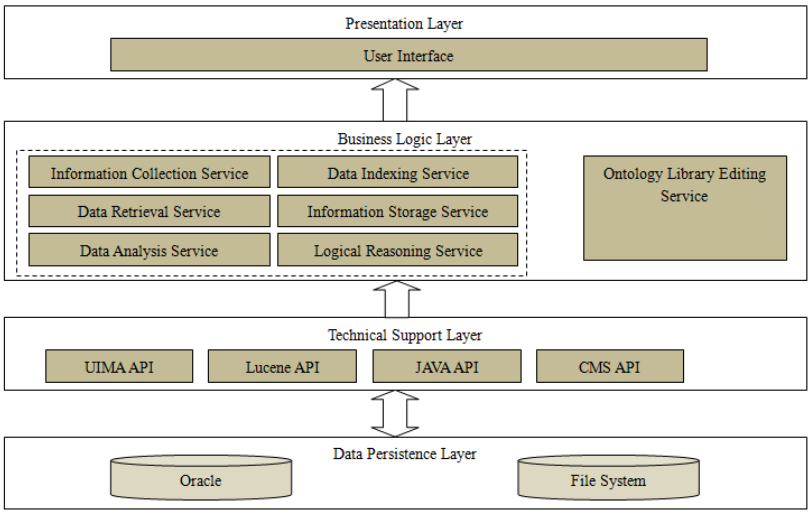

Figure 1. Four-layer Technical Framework of LASP

LASP can be divided into four subsystems according to the functions offered, namely unstructured information content management subsystem, unstructured information analysis subsystem, content indexing subsystem and associated retrieval subsystem as shown in Figure 2. Each subsystem completes independent tasks while they also cooperate with each other so as to provide relevant service for legal affairs staff.

A brief introduction to the functions and modules of LASP subsystems are as follows:

(1) Unstructured Information Content Management Subsystem

This subsystem is responsible for the integration of data of heterology and isomerism. The data source integration module integrates and collects heterogenous data in different servers, conduct discriminatory analysis of isomerous unstructured data content therein by virtue of isomerous data analytic technique, and then store such content into CMS system through the Bridge module.

(2) Unstructured Information Analysis Subsystem

This subsystem acquires information from the unstructured information content management subsystem and conducts text analysis to the content through integrated CMS access interface in the way of asynchronous access. Text analysis is a process when this subsystem conducts meaning extraction and is carried out in three steps namely testing, analyzing and tabbing. UIMA provides standard input and output formats and expandable interpreters for this process in order to meet different business requirements.

(3) Content Indexing Subsystem

This subsystem conducts text segmentation of contents processed by the unstructured information analysis subsystem in the way of synchronous access and incorporates such content into the indexing set after the processing of the content indexing module.

(4) Associated Retrieval Subsystem

By building the ontology library and the semantic reasoning module, this subsystem conducts ontology based semantic reasoning and retrieval, visits the indexing set and acquires eligible data recording through the full-text retrieval module.

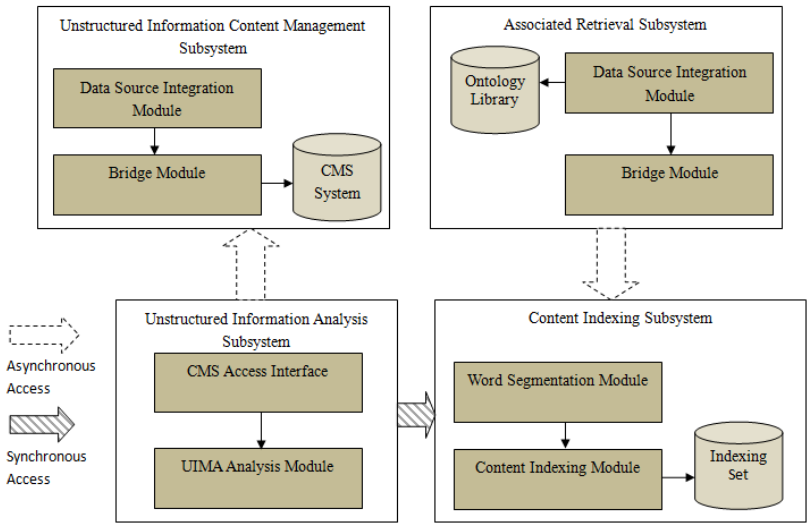

Figure 2. LASP System Structure

5. Application of Big Data Technology in the Information Service Platform

Big data, or mass data, refer to data too big to be intercepted, managed, processed and cleared up into understandable information by manual work within reasonable time. Given that the total data volume is the same, comparing to separate analysis of independent small data sets, integrating such small data sets before analyzing usually generates additional information and data relations which can be employed to detect business trend, judge research quality, avoid spread of diseases, fight crimes and measure real-time traffic condition; such purposes are exactly the reason for the prevalence of big data sets. Since such platform is to manage, analyze and inquire a large number of unstructured data, the application of big data technology will facilitate the application of the system. As cloud computing is the core of big data analyzing and processing technology as well as the basic platform for the analytical application of big data, all big data processing technologies and application platforms within this information platform are based on cloud computing, most typical examples being UFS - a distributed file system, MapReduce - a batch processing technology and BigTable - a distributed database.

a. Distributed File System

Unlike the case of structured data, it's troublesome to present unstructured data including office documents, texts, images, videos and audios, all in different formats, by database two-dimensional logics, and therefore distributed file system is employed as the key technology for unstructured data storage. This platform utilizes GFS (Google File System) as reference and its system framework is shown in Figure 3.

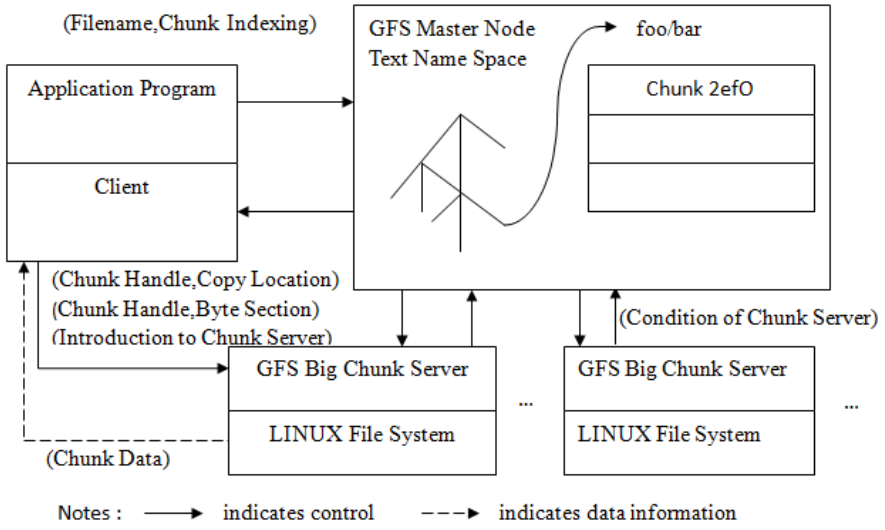

Figure 3. GFS System Framework

The system consists of three parts: Client, Master and Chunk Server.

(1) Client is a set of special access interfaces provided by GFS in the form of library files for application programs regardless of POSIX standard. Application programs invoke these library functions directly and get linked with such library;
(2) Master is the management node of GFS and it mainly stores data file related metadata rather than data chunk. Metadata include name space, i.e., the directory structure of the whole file system. Besides, Master node receives updates from each Chunk node periodically to guarantee the time-effectiveness of metadata;
(3) Chunk Server is responsible for integral storage Chunks.

b. Distributed Parallel Database

We can see that original data acquired from data source are stored in distributed file system, but users get used to accessing files from the database; traditional relation-based distributed database has failed to meet data storage requirements in a big data era. Therefore, this platform employs BigTable, a database system program published by Google. The program provided users with a simple data model which mainly takes advantage of a multidimensional data table where inquiry and positioning are conducted by virtue of row keyword, column keyword and timestamp and users can control data distribution and format by themselves in a dynamic manner. The basic framework of BigTable is shown in Figure 4. Data in BigTable is stored in subtable server in the form of subtables, master server sets up subtables and stores data in GFS file system in the form of GFS; meanwhile, the client communicates directly with subtable server, and Chubby server conducts condition monitoring to subtable server; master server can check Chubby server to observe the condition of subtables, shut down the faulted servers in case of abnormality and transfer its tasks to other servers.

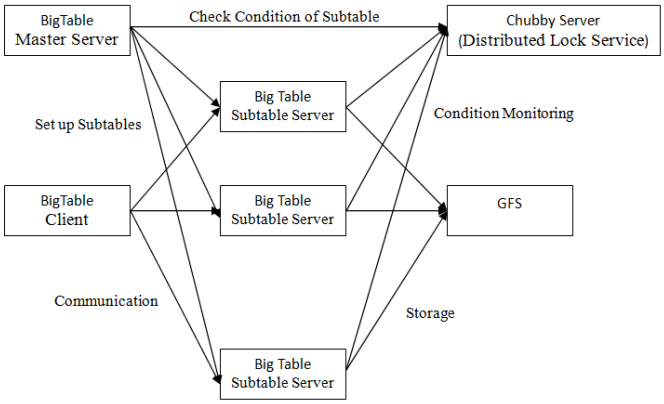

Figure 4. Big Table Basic Framework

6. UIMA and its Application

a. Brief Introduction

Traditional knowledge discovery technology mainly involves structured database, and keywords are generally used to search information in traditional search engines, e.g. if you input "legal malpractice", the search engine will return all webpages containing "legal malpractice". However, such results can obviously not meet actual search requirements since there are no uniform management to and structure of unstructured information resources, leading to the ignorance of abundant connotative knowledge. A more intelligent search engine is desired which can return webpages containing the classification of specific legal malpractices such as "crime of abuse of authority", "crime of dereliction" and "crime of leaking state secrets intentionally", as well as audios and videos of corresponding cases and the results. This is called semantic search technology through which users can get satisfactory answers after inputting questions using everyday language. It can "understand" user's language and give the most precise answer to their questions.

Unstructured Information Management Architecture (UIMA) is the system architecture of unstructured information management. Cases, evidences and testimonies in legal affairs are not in the form of text, making it challenging to dig out associated information from such non-text information. Through UIMA, unstructured information such as audios, videos and images can be transformed into structured information and stored into database and index. It can greatly save attention on irrelevant information and improve information processing ability and data mining efficiency. See Figure 5.

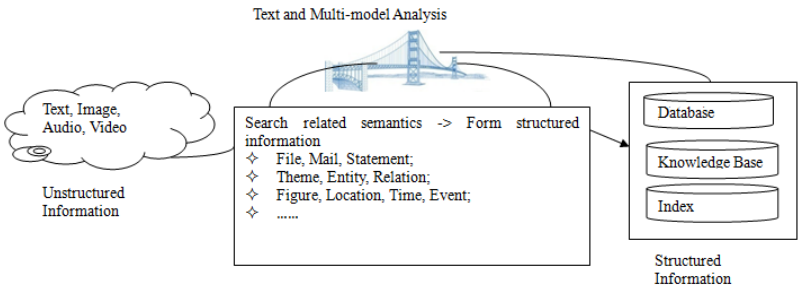

Figure 5. UIMA's Role as a Bridge between Structured Information and Unstructured Information

b. Structural Analysis Of UIMA

According to the standard definition of UIMA, it mainly consists of seven components, namely, Common Analysis Structure (CAS), Type System Model, Base Type System, Abstract Interfaces, Behavioral Metadata, Processing Element Metadata and Services WSDL Descriptions.

(1) Common Analysis Structure

Common Analysis Structure (CAS) is the core data structure to conduct data abstraction to the Analytic (Artifact) as well as the data channel for data analysis and intercommunication of UIMA components.

CAS is the internal structure of object graphs which contains object instances generated according to the types in the type system. In CAS, the association between object graph and type system is described by Slot relation. Slot refers to a set of (feature, value) key value corresponding relations.

(2) Type System Model

Type System Model defines the data model in actual application scenario and is realized by developers according to type system interfaces. A Type defines the object, attribute, method and a series of constraint conditions according to the characteristics of actual business.

(3) Base Type System

Base Type System is a basic type system defined by UIMA that provides lower-layer support for Type System Model.

Base Type System offers eight metadata types: EString, EBoolean, EByte (8 bits), EShort (16 bits), EInt (32 bits), ELong (64 bits), EFloat (32 bits) and EDouble (64 bits).

Besides, Base Type System offers Sofa (Subject of Analysis) and Annotation type system for the storage of the Analytic (Artifact) and annotations. The analytic refers to an information unit used for analysis, while annotation refers to valuable contents in the analytic.

(4) UIMA Abstract Interface

As shown in Figure 6, abstract interfaces specified by UIMA provide definitions to standard components and procedures and process unstructured data information through the processing of CAS. UIMA Abstract Interface specifications define two subtypes developing from Processing Element (PE): the Analytic and FlowController.

Processing Element interface defines two methods, namely, getMetadata and setConfigurationParameters. All Processing Element service components should implement this interface.

Analytic interface is used to analyze the content of CAS and it has two subtypes: Analyzer and CasMultiplier. Analyzer interface updates the input CAS while CasMultiplier interface divides the input CAS into two or more CASs as output or combines two or more input CASs into one CAS as output.

As the workflow controller of analysis, FlowController interface determines the choice of analysis paths among multiple Analytics.

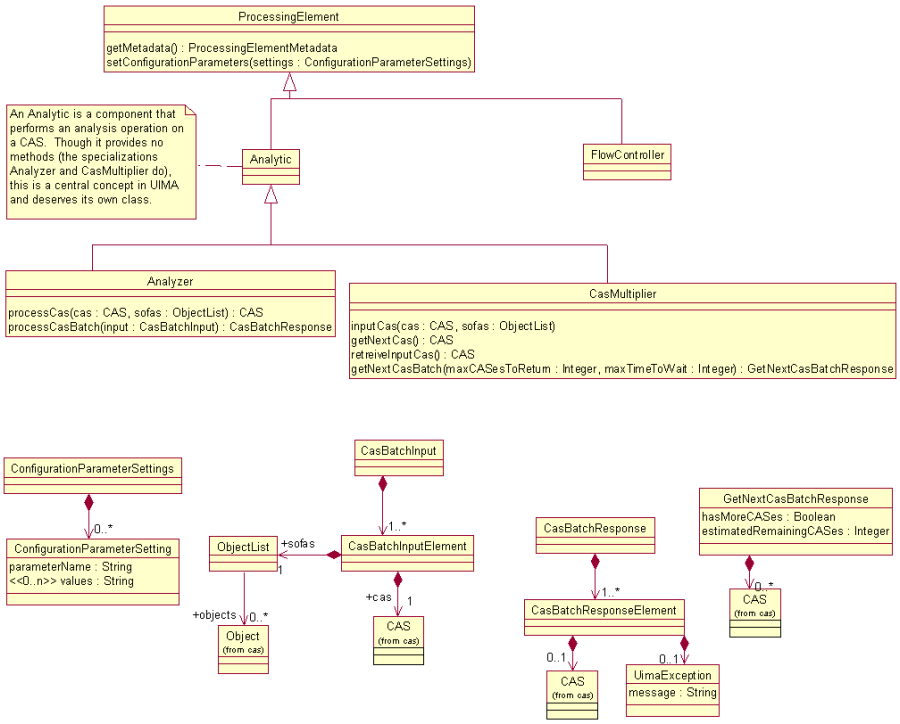

Figure 6. UIMA Abstract Interface Specifications

7. Word Segmentation and its Application

Chinese word segmentation means dividing a Chinese character sequence into separate words. Word segmentation is a process where a consecutive character sequence recombines into a word sequence according to certain specifications. In the writing of English, spaces are the natural separators between words; in the writing of Chinese, characters, sentences and sections can be simply divided by obvious separators while there is no formal separator for words. Chinese word segmentation is the basis for text mining. As for a section of Chinese, successful word segmentation helps the computer identify sentence meaning automatically. The most important function of a search engine is not finding all results since no one can read all results from tens of billions of webpages, but listing the most related results up front, which is also called relevancy ranking. The accuracy of Chinese word segmentation has direct impact on the relevancy ranking of search results. Common word segmentation algorithms are mainly: maximum matching with backtracking, reverse maximum matching with backtracking, omni-segmentation method and maximum probability segmentation method.

Maximum matching with backtracking (MMB) aims to find out the longest word. It firstly supposes that the length of the longest word is L, gets a word string in the length of L starting from the first character of the sentence and performs matching. If the matching is successful, it then regards such word string as one word and repeats the above process from the following character; if the matching fails, the last character of the word string is taken out and matching is performed again until the matching is successful or the word string is empty. The algorithm of reverse maximum matching with backtracking (RMMB) is the same with that of MMB except that segmentation is performed from the right to the left and in case of matching failure the first character is taken out. It is equal to reversing the word string to be segmented character by character to form a new word string and then performing MMB. RMMB and MMB have their blind spots in segmentation and they are unable to spot all overlap ambiguities. Omni-

segmentation method can get all reduction formulas of the given word string through omni-reduction process, i.e. it can get all feasible segmentation solutions and eliminate blind spots. The basic idea of maximum probability segmentation method is that a Chinese word string to be analyzed may have more than one segmentation results and the word string with the biggest probability is adopted as the result. There are many Chinese word segmentation tools at present and commonly used Lucene-based ones are Paoding Analysis, Imdict-chinese-analyzer, mmseg4j, IKAnalyzer, etc.

This paper performs Chinese word segmentation under the standard of IKAnalyzer 3.x which is an open-source lightweight-class Chinese word segmentation toolkit developed based on java language. It adopts the exclusive "forward iteration finest grit segmentation algorithm" with a high-speed processing capacity of 600,000 characters per second. A multiple-subprocessor analysis model is adopted to support the segmentation of English alphabets (IP address, Email and URL), numbers (date, commonly used Chinese quantifier, Roman numerals and scientific notation) and Chinese vocabulary (person name and place name). It allows optimized dictionary storage and smaller memory usage and supports user dictionary extended definition.

The framework of IKAnalyzer 3.x is shown in Figure 7. The most important APIs of IKAnalyzer are as follows:

IKAnalyzer: the main type of IK word segmentation tools, and also the Lucene Analyzer type implementer of IK word segmentation tools.

IKQueryParser: an optimized query analyzer directing at Lucene full-text retrieval.

IKSimilarity: the similarity evaluator of IKAnalyzer. It reloads the Coord method of DefaultSimilarity and increases the weight of lemma hit number in similarity comparison. Document similarity will rise when two or more lemmas are matched.

IKSegmentation: the core type of IK word segmentation tools. It is also the implementer of analyzer in its true sense.

Lexeme: the semantic unit object of IK word segmentation tools equivalent to Token lemma object of Lucene.

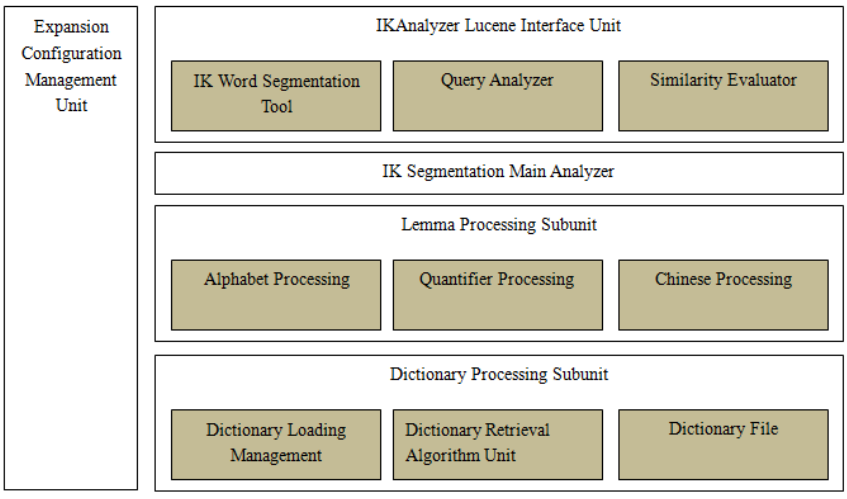

Figure 7. IKAnalyzer 3.x Framework

8. Conclusion

With the increasing informatization grade of the legal industry itself, more and more services in legal affairs are offered in information-based forms, and there is a large number of unstructured information in such business data. It's a problem the legal industry needs to tackle in information management to rapidly acquire valuable contents from the massive unstructured data and make use of such contents. Although

traditional full-text retrieval technology enables quick search of matching materials according to keywords, there are many problems. This paper proposes the solution of an information-based platform by taking advantage of cloud computing platform on the basis of UIMA specifications and full-text retrieval technology. It firstly integrates data sources of unstructured information of heterology and isomerism in the legal industry, and carries out uniform management of information resources by means of Content Management System (CMS). Then expandable UIMA framework is employed to conduct data acquisition and expanded data analysis of such unstructured legal affairs information resources, and Lucene indexing technology is utilized to conduct sequential indexing of data content and analysis results. To guarantee more scientific and accurate retrieval, this paper adopts IKAnalyzer 3.x technical framework and designs a solution of Chinese word segmentation. Through this platform, legal affairs enterprises can effectively integrate structured and unstructured information resources and realize the storage, analysis, retrieval and decision-making applications of business data content.

Acknowledgement

This work is financially supported by Scientific Research Project of Hebei Science and Technology Department, China (No. 13215325)

References

[1] Q. Lele and Z. Xin, "Design and Realization of Information Service Platform of Logistics Parks Based on Cloud Computing", AISS, vol. 4, no. 23, **(2012)**, pp. 112-120.
[2] Z. Guigang, L. Chao and Z. Yong, "A Kind of Cloud Storage Model Research Based on Massive Information Processing", Journal of Computer Research and Development, vol. 5, no. 9, **(2012)**, pp. 32-36.
[3] J. Baihua and S. Hansheng, "Research on hazardous chemicals logistics platform based on IOT", Computers and Applied Chemistry, vol. 31, no. 10, **(2014),** pp. 1271-1274.
[4] W. Chen and Q. Lele, "Design and Realization of Legal Affairs Information Query & Analysis System Based on Cloud Computing & UIMA", AISS, vol. 4, no. 23, **(2012)**, pp. 189-197.
[5] Z. Yuanzhong and J. Minzheng, "Research on Safety Supervision of Dangerous Chemical Enterprises Based on the Internet of ThingsTechnology", Journal of Beijing Polytechnic College, vol. 4, no.1**, (2015),** pp. 1-6**.**
[6] Q. Lele and L. Wei, "Key technology research in examination integrated management information system", Hebei Journal of Industrial Science and Technology, vol. 27, no. 4, **(2010),** pp. 213-215, 235**.**
[7] Z. Guigang, L. Chao and Z. Yong, "A Kind of Cloud Storage Model Research Based on Massive Information Processing", Journal of Computer Research and Development, vol. 5, no. 8, **(2012)**, pp. 32-36.
[8] D. Ferrucci and A. Lally, "Building an example application with UIMA", IBM System Journal, vol, 11, no.6, **(2004)**, pp. 105-121**.**
[9] F. Guohe and Z. Wei, "Review of Chinese Automatic Word Segmentation", Library and Information Service, vol. 9, no. 4, **(2012)**, pp. 41-45.

Authors

Wang Chen, she was born in Shijiazhuang, Hebei, China on May 23, 1981. She earned a Master's Degree of Engineering from Hebei University of Economics & Business, China in 2007. Her major field of study is civil and commercial law, intellectual property law and information system analysis of law.

She is working in the Office of Scientific Research in Hebei University of Science & Technology. She has been teaching more than 10 courses at various levels including Law Management Information System and Intellectual Property. She has published more than 10 articles, of which 4 papers were indexed by EI.

