

## DESIGN SPECIFICS FOR A LEGAL INFORMATION RETRIEVAL SYSTEM OVER WAP

Vanyo G. Peychev

The article discusses the problems of a legal information system design that uses the mobile net as a transfer environment. Some approaches for a legal database organization are described and a prototype design is presented. The solution can be used for developing information-searching systems in an environment with limited resources.

The information retrieval [1] is an interesting task which solution eases the people in the information searching process. The absence of a universal description of the document structure is a reason to look for a partial solution. This solution ensures the automated processing of documents with predefined structure [2]. Legal documents are a realistic example of such documents that have a relatively stable structure.

The aim of the article is to design a database prototype of legal documents to support search over a Wireless Application Protocol (WAP). WAP imposes strong limits over data exchanged. The database prototype is designed for better search satisfying specific user requirements.

**Analysis of the system requirements.** The analysis process includes research on the specifics of the problem. A specification of the users' requirements is built as a result of this process. Information for legal documents, law practice, court decision – that is the common scope of the legal information systems. But some of them are specialized into a certain legal area [3]. Thus, the used databases volumes are different and they depend on the information type we want to manage. On the other hand, when we want to manage data through WAP, we should remove all redundant data and additional information, if it exists. When we talk about legal documents, we should define a structure for a legal document in order to clear unnecessary information.

The users can think for the legal information systems as an additional feature to the traditional sheet-based information bearers. Two directions are important for the information retrieval process:

- the information gathering and
- the information searching.

For the first one the important tasks are the choice of an information structure and how the information will be saved.

The full texts of legal documents and court decisions are the preferred type of the stored information. Regarding the information details, the users prefer to get information

about individual documents [4]. When legal documents are concerned, the users are most likely to be given detail information, for example separate part of the document.

For legal documents the searching strategies trend is searching by keywords and by properties of the documents. When the users define a search question they use additional resources to specify the keywords.

When we define a question using the language of keywords the logical operators “AND”, “OR” and “NOT” are evaluated as particularly important. The questions are defined with simple conditions. The preferred strategy is a sequence of simple questions versus a single big and difficult task. The idea is to follow a given and simple strategy rather than a long processing with uncontrolled intermediate results.

An important condition for the information systems design is how the users will interact with the system. Depending on their professional qualification, the users of legal information systems can be classified as:

- Users for whom the usage of legal information systems is directly associated with their professional activities;
- Users for whom the usage of legal information systems is a chance occurrence. Their information necessities have arisen in relation with a concrete event or outside activities.

For these two types of users of legal information systems the major activity is searching for legal content into documents.

**WAP features.** WAP is an open standard for communication between a mobile device and computer networks. WAP is optimized to work with limitations enforced by the environment and the mobile devices [5]. Some of them are: a small display, very little memory and a narrow bandwidth for data carrying.

As there are only few rows and columns onto the display, the data input and output is limited. A good WAP site should present to the user little but reasonable information. The introduction of simple ways for navigation is another possibility for improving a WAP site. Solving these two problems, WAP enables the design of a convenient visual environment for information stream management. At the same time this information becomes accessible by mobile devices. The limited resources determine the necessity of a fast and easy access to the information.

For this reason a key point into building a WAP access to information content is a usable visual interface. If we divide the information to the level of a grammatical paragraph, the number of hyperlinks for the document will increase significantly. The normal practice is one document to be fitted into one page as it is done for most HTML documents. But even in the HTTP (HTML) world it is not possible or not reasonable for every document to be displayed into one page – ex. the list of Bulgarian customs tariffs. Under these circumstances the user navigation in a model for WAP based hypertext system would become more complicated. This could lead to loss of orientation and comfort.

**Special features for a legal database design.** The ensuring of appropriate information as a result of a searching process into legal documents is concerned with:

- Pre-definition of the documents structure (1)
- Development of strategies and methods for finding of information content (2)

These aspects are very important when we design the search into legal documents

through WAP. The result information, when we search into legal documents, is assumed as generalized or detailed. The generalized information is the information which by the means of hyperlink gives us the possibility to access the detailed information. Detailed information is the information which the users receive and can use.

The legal information systems manage data which are object of the law science. The base object in the law science is legal document, for example legal acts and court decisions. The essence of these objects is the grammatical sentence. It defines the type of data which should be saved in every legal information system.

Thus, structuring of the data is significant. The following strategies can be used for data saving:

- the data are saved as grammatical paragraphs;
- the data are saved as words and rules.

The first approach (1) supposes the storing of a minimal legal construction, which is logically connected text. Usually these constructions in legal documents are elements of the document, for example parts, articles, points, etc. The advantages of this approach are connected with:

- the introducing of new information into the database;
- a lack of grammatical rules to construct the answers. The text that is found is presented directly to the user without any additional processing;
- easy ways for adding of renewals into the document;
- the stored sentences can be accessed without using a specialized information system.

The disadvantages for storing data as grammatical paragraphs are as follow:

- we cannot define a hyperlink;
- methods for searching are limited only to a template match into the range of the stored grammatical construction;
- the amount of repeating information in the database increases.

The later approach (2) includes a definition of a data dictionary and grammatical rules to construct the sentences. The storing of single separate words from the document into the database requires a predefined strategy how to construct grammatical sentences. This strategy can be also stored into the database.

The advantages when we save data dictionary and rules are:

- hyperlinks can be defined for separate word(s) into a grammatical sentence;
- search strategies and methods are simplified. They enable the construction of simplified sentences. This makes the searching process faster;
- the amount of repeating information in the database decreases.

The disadvantages of grammatical saved constructions are:

- the introducing of new information into the database can be a serious problem. To solve this problem an information system should be created to handle the input necessities;
- there is no way to extract information without using of the stored rules.

For legal information retrieval systems over WAP we consider the data to be stored as data dictionary and rules. The aim is to make an efficient search in legal documents and to present the result using WML.

**Requirements to design a legal database for WAP networks.** Depending on the information completeness, the databases for legal information systems [6] are

separated into:

- report-oriented databases;
- databases that contain the full texts of the documents.

The report oriented database redirects the user to the publisher or to a place where the document is published. For some types of information, for example legal literature, a reference to the source is saved. Databases that contain the full texts of the documents present to the user a part or the whole text of the document.

Depending on the application, the databases for legal information systems are classified as:

- general;
- specialized.

The legal databases with a general application have the following typical extent:

- legal acts – laws, decisions, declarations;
- court practice;
- legal literature.

They are usually supported by government offices. Their aim is the coherent storage and support of the actual state for the given information types. Specialized database are created for specific legal documents. They also store information about documents, but they differ from the general databases by the methods for information retrieving and convenient use.

The general activity that is supported by a legal information system is the search in a legal document. The article presents a database structure, which is appropriate for user interaction over WAP. To support a good user interface over WAP, the database should be designed as a database with the support of the full text of documents. And the information system should be specialized. The major aim is the users to obtain the searched information in the whole content without an additional collecting of information from another source. For achieving a successful result the following tasks are defined:

- create a database – it will improve the information searching process. The expected result is a minimal-volume grammatical construction;
- create hyperlinks – for separated words and/or phrases – and store the hyperlinks into the database;
- store a time structure for legal documents – this is similar to store different versions of the document. The validity time of every paragraph will be stored, too.

**Database prototype design.** To solve defined tasks the following objects are considered (Fig. 1):

- legal document;
- grammatical sentence (the text of elements of a legal document);
- types of elements in a legal document;
- dictionary;
- hyperlinks.

Every legal document has properties that are unique for the document. These properties are as follows:

- document header – the name of the document;
- creation date of the document;

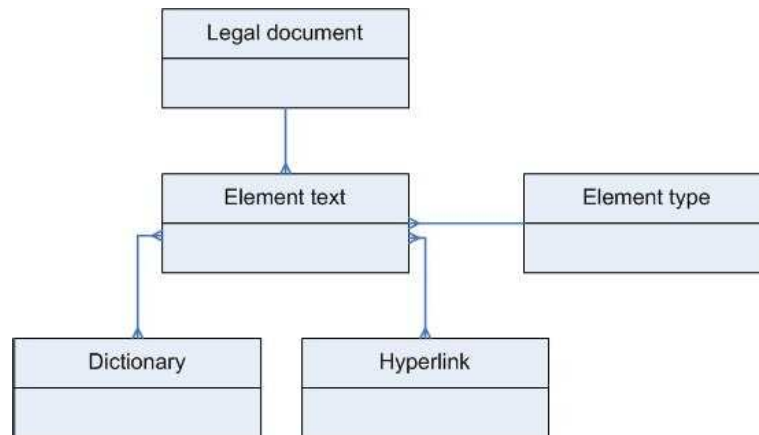


Fig. 1. Schema of object relations

- status of the document – actual state of the document. Possible values are – valid, invalid and changed.

Every document consists of elements which are specific for its structure. Every element presents a grammatical sentence or set of sentences. The basic properties of this object are as follows:

- text of element;
- date, when the element is changed.

Every element from a document has a type. The type is defined from the legal document structure. This object should be considered as a nomenclature object. Its basic property is the description of the type of an element. This object is open and new type of elements could be added. This means that the set of the supported documents can be increased.

Dictionary is the object which contains the used words and punctuations. Its purpose is to save all used words and punctuations, which are entered into a given legal document. This object could be considered as a nomenclature object, too.

The object that saves all hyperlink relations for the legal document is called hyperlink. Similarly as the two objects above, the hyperlink object is considered as a nomenclature object. This allows a new hyperlink to be added for the existing document.

Different types of documents can be supported by the addition of their specific element types in the database.

**Conclusion.** The solution points to some interesting problems of creating a database for an information retrieval over WAP. The main achievements are:

- design of a database for legal documents – the database design is oriented to fast searching;
- creating and saving hyperlinks for separated words and/or phrases into the database;
- support of a time structure for a legal document;
- generating information content from a legal document appropriate for a mobile device with resource limitations.

The database prototype we discussed can be used as a model to create information retrieval systems for an environment of resource limited devices.

## REFERENCES

- [1] C. RIJSBERGEN. Information retrieval. Department of Computing Science, University of Glasgow.
- [2] V. PEYCHEV. XML model for legal documents. *Problems of Engineering Cybernetics and Robotics*. Bulgarian Academy of Science, BAS, Sofia, 2004, vol. **54**, 86–92.
- [3] M. MOENS. What Information Retrieval Can Learn from Case-Based Reasoning, Legal Knowledge and Information Systems. The Fifteenth Annual Conference, Amsterdam: IOS Press, 2002, 83–91.
- [4] A. BOLIOLI, L. DINI, P. MERCATALI, F. ROMANO. For the Automated Mark-Up of Italian Legislative Texts in XML, Legal Knowledge and Information Systems. The Fifteenth Annual Conference, Amsterdam: IOS Press, 2002, 21–30.
- [5] R. HOWELL. WAP Overview. Wrox Conference, 2000.
- [6] A. MUNTJEWERFF. Principled and Structured Design of Electronic Materials for Learning the Law, Legal Knowledge and Information Systems. The Fifteenth Annual Conference, Amsterdam: IOS Press, 2002, 133–142.

Department of Computer Sciences  
Faculty of Mathematics and Informatics  
University of Sofia  
5, James Baucher Str.  
1164 Sofia, Bulgaria  
e-mail: vanyo@fmi.uni-sofia.bg

## ОСОБЕНОСТИ НА СЪЗДАВАНЕ НА ПРАВНА ИНФОРМАЦИОННА СИСТЕМА ПРЕЗ WAP ПРОТОКОЛ

**Ваньо Г. Пейчев**

В настоящият доклад се дискутират проблемите при създаване на правна информационна система чрез използване на мобилна мрежа (WAP) като преносна среда. Дискутират се възможните начини за организиране на правна база от данни. Представено е примерно решение за създаване на правна база от данни. Решението може да се използва като основа за разработване на информационно-търсещи системи в среда с ограничени възможности на крайните устройства.