

CITEXPERT - An Expertsystem for Automatically Recognizing Legal Referenes and Generating Hyperlinks

Felix Gantner
infolex
Rotgasse 2/7
1010 Wien
Tel: ++ 43 1 535 69 41
E-mail: gantner@infolex.co.at

Abstract:

Citation of legal authorities are an important part of legal documents. They can be used for improving the usability of legal databases and hypermedia-systems by working up the references into hyperlinks and by extracting metainformation about the content of the legal text. CITEXPERT is an expertsystem for automatically recognizing references to statutes in a legal text and for generating metainformation about and/or inserting hyperlinks into the legal text.

In this system pattern recognition using regular expressions is used for searching for legal references. The recognized patterns are analysed by rules depending on the different types of references.

For assigning the found patterns to the legal rules referred to, a thesaurus of the names of statutes and commonly used short forms and terms of referring was built up. CITEXPERT is used for working up decisions of the Austrian Supreme Courts for legal databases.

1. Introduction

References to legal authorities and precedents are an essential part of legal documents. Almost every legal text contains some citation of legal authorities. In legal databases and hypermedia-systems this citation of authorities can be used to improve the usability of these system. In hypermedia-systems the user-friendliness can be improved by working up the citation of authorities into hyperlinks to other (legal) documents the system contains. So the user can take advantage of the complex network of information described by the references in the documents for investigating legal information [1]. Besides, citation of authorities can be used to extract metainformation about the the legal text too. Especially the references to statutes a document contains can be used for generating an approximate description of its legal content. If standardized descriptors for legal rules are used in this description, the statutes referenced in the document can be used for efficient investigation in legal full-text-databases. Since there is no uniform way of referencing to legal rules, only standardized descriptors make it possible to search for

specific statutes in a database more effectively and to receive a complete result.

Since working up legal references manually is a big effort, there are several attempts to build systems for analyzing the content of legal text automatically and generating hyperlinks or metainformation ([2], [3]).

In this paper CITEXPERT, an expertsystem for recognizing references to statutes in legal text, is described. It is used for working up decisions of the Austrian Supreme Court, the Austrian Constitutional Court and the Austrian Administrative Court.

2. Preliminaries

CITEXPERT is a system for the automatic recognition of references to statutes and for generating hyperlinks to or standardized metainformation about these statutes.

There are two kinds of references to statutes CITEXPERT recognizes:

- citation referencing to a complete subdivision of a legal rule
- citation without subdivision referring to a source of law

2.1. References to complete subdivisions

A citation referencing to a complete subdivision of a legal rule in general consists of the complete subdivision and of the name of the statute.

In general the structure of a citation is:

Subdivision + Name of the Statute

with:

Subdivision = (tag + numbering)_{first level} + (tag + numbering)_{lower levels}

A subdivision is complete if it starts with a first level subdivision.

Examples for references to complete subdivision are:

- a) § 1358 ABGB
- b) § 2 Abs 1 Tiroler FischereiG 1952
- c) Art. 7 B-VG
- d) Art XXVIII Z 2 EO-Novelle 1991, BGBl 1991/628
- e) § 14 Abs 1 zweiter Satz EO
- f) § 2 Z 11 des Art 1 FBG
- g) §§ 519 Abs 1 Z 1 und 528 Abs 2 Z 2 zweiter Fall ZPO
- h) § 2 Z 1-3, § 3 Z 1 AnFO
- i) §§ 158c bis 158h VersVG
- j) § 33 TP 21 Abs 1 Z 2 GebG
- k) TP 3 A I 1 lit.d RAT

The examples a) to e) show the most common structure of a reference to a legal rule in decisions of the Austrian Supreme Court („the standard way of referencing“). Only one legal rule is referenced and the subdivision may contain only one first level subdivision (a, c) or two (b, d) and even more (e, j, k) levels of subdivision. The source of law may be a commonly used short form of the name of the statute or a longer description, which may include the organ, where the statute was published, too.

As examples j) and k) show, some of the tags marking first level subdivision like „§“ or „Art“ may also be found in lower levels of subdivision. Some other tags like „Abs“ only mark lower levels of subdivisions and must not be found in first level subdivisions.

Tags marking a subdivision are for instance § (= section), §§, Artikel (= article), Absatz (= paragraph), Punkt (= point), Ziffer (= Number), Anhang (= appendix) ...

For all tags different spellings have to be considered. So the tag „Artikel“ can be found in legal documents as „Art“, „Art.“, „Artt“, „Artt.“, „Artikel“.

The numbering may contain numbers, letters or roman numbers in any sequence.

Besides, another way of referencing like in f) is found too.

The structure of these references is:

Subdivision_{lower levels} + „des“ (= of) + Subdivision_{first level}

Subdivision_{lower levels} + Name of the Statute

The recognized subdivisions of references of this kind have to be brought into the order of the standard way of referencing.

If more than one legal rule is referred to in a citation, there are two ways of forming a reference.

The first is a list of subdivisions followed by the name of a statute. Examples for this kind of citation are g) and h).

Important for this kind of references is, that frequently the tags in the first level are omitted if the precedent subdivisions is the same level (example g)). So the tag and the level of the subdivision can only be figured out by taking into account the context of the reference.

It may happen too that even those parts of the subdivision which are identical with the precedent one are omitted.

Most references of this kind can be interpreted correctly.

However in some cases the level of the subsequent subdivision cannot be found out unambiguously.

So within subdivisions like § 2 Abs 1 und 3 the numbering 3 can be interpreted as § 3 or as part of the subdivision § 2 Abs 3.

If legal rules with numbering that follows one another are referenced, references (like examples h) and i)) are often used for describing a range of rules. The beginning and ending rule are connected by a term like „bis“ (= to) or „-“. When references to a range of legal rules are recognized and worked up into meta-information or hyperlinks, all the individual rules of the range should be taken into account.

2.2. References containing sources of law without subdivision

The second kind of citation are references to legal rules containing sources of law like the federal gazette (Bundesgesetzblatt, BGBl), gazettes of provinces (Landesgesetzblatt, LGBl) or the gazette of the EC (ABl). Examples for references of that kind are:

- l) Richtlinie 93/42/EWG des Rates vom 14.6.1993
- m) Bundesgesetz vom 14.2.1973 BGBl Nr.108
- n) Auslegungsrichtlinie 88/361/EWG, ABl L Nr 178 vom 8.7.1988, Seite 5
- o) BGBl 320/1988

These references can contain in addition to the source of law the name of the statute, date of publishing or a formalized description of the legal source.

In some cases the name in references with subdivision contains a legal source too (example d)). In spite of referring to a legal source, citation of this kind is interpreted as a first kind one.

In the following we will understand a legal reference as a citation of one of the two kinds described before.

3. CITEXPERT

Legal text is analyzed with CITEXPERT in four steps:

- 1) Preliminary preparation of the legal text
- 2) Pattern recognition and rule-based analysis

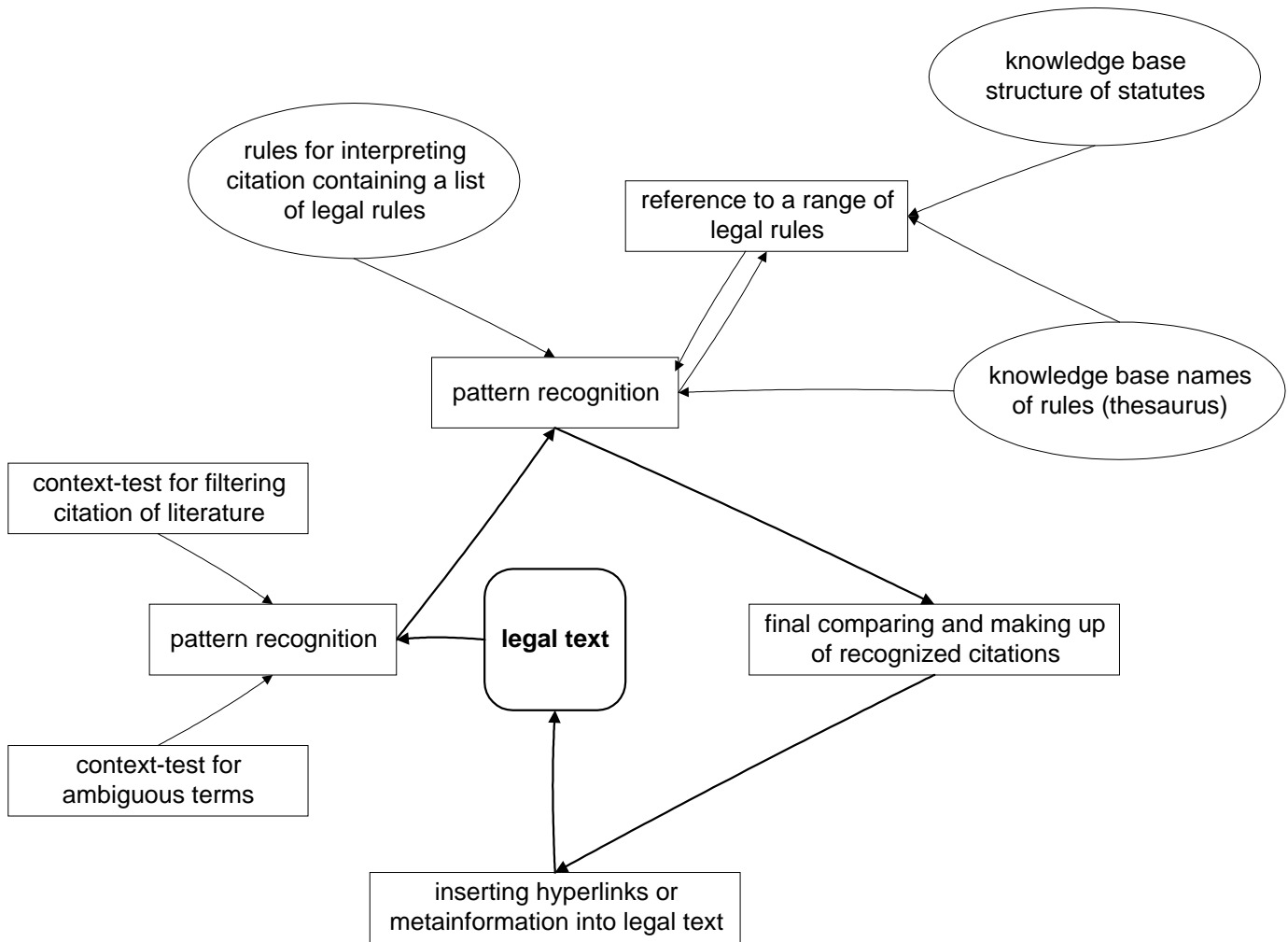
- 3) Final comparing and working up of recognized citations
- 4) Adding hyperlinks or metainformation to the legal text

3.1. Preliminary preparation

In this first step the legal text is prepared for the recognition of citation of legal rules. A temporal copy of the document is divided into paragraphs and all characters for formatting the text are removed.

Since CITEXPERT is used for working up documents for a database containing decisions and statutes, only references to statutes are used for generating metainformation about the documents. Citation of literature is removed from the temporal copy of the document to make sure that the two types of citation do not get mixed up.

Terms which can be mixed up with tags of subdivisions of legal rules because of the ambiguity of words are looked up too and after an analysis of the context removed too.



CITEXPERT

After this the paragraphs are searched for references to literature like commentaries (eg „MGA ZPO14 § 25/11“), which also may refer to legal rules. Although these references contain tags and numbering of legal rules similar to citation of statutes, the structure, context and some tags like references to margins are different. So a specific set of rules for pattern recognition can be defined for localizing citation of literature.

3.2. Pattern recognition and rule-based analysis

After this preliminary preparation the text is searched for legal references.

For every kind of citation an extensive set of rules has been defined. These rules describe all kind of tags used in legal references, all different spellings of the tags and all the

kinds of numbering used in subdivisions. The valid sequences of subdivisions are defined as well. Some of these rules could only be defined by analyzing a large number of references, other ones are common knowledge like the structure of roman numbers. These rules were used for describing a finite-state machine with regular expressions ([4], [5]) for recognizing subdivisions of references to legal rules. For those references containing sources of law without subdivision a set of rules and regular expressions were defined too. Using these regular expressions the references to legal rules in the text are recognized. So the subdivisions and the references containing sources of law without subdivision are found. The references containing sources of law are compared with

BG BGBI 1980/320
BG vom 2. 7. 1980, mit dem das Gesetz über Gesellschaften mit beschränkter Haftung geändert wurde (BGBI. 320)
BGBI 1980/320
BGBI.Nr.320/1980
Bundesgesetz BGBI Nr 320/1980
Bundesgesetz über die Änderung des Gesetz über Gesellschaften mit beschränkter Haftung vom 2 Juli 1980
Bundesgesetz vom 2 Juli 1980, BGBI Nr 320
Bundesgesetzes BGBI. Nr. 320/1980
Bundesgesetzes vom 2. Juli 1980, BGBI. Nr. 320
Bundesgesetzes vom 2. Juli 1980, mit dem das Gesetz über Gesellschaften mit beschränkter Haftung geändert wird, BGBI.Nr. 320
Gesetz vom 2 Juli 1980, BGBI Nr 320
GmbH-Gesetzovelle 1980
GmbH-Nov. 1980 BGBI. 320
GmbHNov 1980
Nov. BGBI. 320/1980
Novelle BGBI.Nr.320/1980

Some of the entries in the thesaurus for one statute

a thesaurus which contains all the the names of statutes and all commonly used short forms and terms of referring. Yet unknown terms are recorded, so the thesaurus can be improved regularly. Found subdivisions are tested wether they are part of a list of subdivisions in a citation of several legal rules. If they are, the next subdivision is looked up and the rules for interpreting the subdivisions and the relation between these parts of legal references are applied. If no more subdivisions are found, the name of the statute they are followed is looked up in the thesaurus. If it is found, it is added to the recognized subdivisions. Unknown names are recorded. After this step the found subdivisions are tested wether they refer to a range of legal rules. If they do so, the structure of the recognized statute is looked up in a knowledge-base,

where the numbering of the subdivisions of the most important statutes are recorded. If the statute is found, all the individual rules of the range can be added to the found references.

3.3. Final comparing and working up

After this step the found references are finally compared and worked up. The tags in the subdivisions of the citation are transferred into a uniform spelling, so that all the legal rules in all the legal documents worked up are referred in the same way. For the law of the provinces some special steps of working up are necessary. Often the province where a statute is applicable is mentioned only in the first citation and later on it is referred without this information. By comparing the

provinces referred to in the decision, this information can be added to all the references. The same goes for years added to the name of the statute. Additionally subdivisions of references which names were not found (for example in case of references, in which only the subdivision is cited and the name of the statute has to be recognized by knowing the context) are compared with other found subdivisions. In case of unambiguous similarities in subdivisions (eg equal subdivisions or levels of subdivisions) names of legal rules can be mapped on each other.

3.4. Adding hyperlinks or metainformation

The last step is adding hyperlinks to the found references or metainformation to the document.

4. The application of CITEXPERT

CITEXPERT is used for working up the decisions of the Austrian Supreme Court, the Austrian Constitutional Court and the Austrian Administrative Court for the full-text CD-ROM-databases of the Austrian State Printing Office. These databases contain the decisions of said courts and the Austrian federal statutes. To every worked up dokument all found legal references are added as metainformation. These added information are

indexed, so that the user of the database can search easily for specific legal rules.

The found references will be used as hyperlinks from the decisions to the statutes the database contains.

Random-tests yield that about 90% of the references in the legal documents are recognized.

The efficiency of the system is about 1000 decisions per hour. Considering the effort for updating and maintaining the thesaurus about 700 decisions/man day can be worked up.

References

[1] SAVOY, J.(1994): Searching Information in Legal Hypertewxt Systems, in: AI & LAW Vol 2, 205

[2] MERKL, D./SCHWEIGHOFER, E. (1997): The Exploration of Legal Text Corpora with Hierarchical Neural Networks: A Guided Tour in Public International Law, in: Intentional Conference on Artificial Intelligence and Law - ICAIL'97, Melbourne, Australia

[3] SCHEITHAUER, D./ SCHWEIGHOFER, E./EHRENTAUT, G./UNTERSMAJR, P. (1997): The Automatic Structuring of Hypertext Links: Statistical Methods, Genetic Algorithms and Classifier Systems, in: Proc 1st AustLII Conference on Computerisation of Law via the Internet, Sydney Australia

[4] FLOYD, R./BEIGEL, R. (1996): The Language of Machines, International Thomson Publishing

[5] FRIEDL, J. (1997): Mastering Regular Expressions, O'Reilly