Intelligent Information Retrieval: KONTERM - Automatic Representation of Context Related Terms within a Knowledge Base for a Legal Expert System

Erich SCHWEIGHOFER

Institute of Public International Law
and International Relations
University of Vienna

Werner WINIWARTER

Institute of Applied Computer Science
and Information Systems
University of Vienna

Abstract

Knowledge acquisition constitutes the bottleneck for the creation of legal expert systems. A certain degree of formalism of legal language is an inevitable prerequisite. Our prototype KONTERM deals with that problem by supporting the process of creating a selective thesaurus for a legal information system which can be used for automatic indexing and document classification. This selectivity is obtained by distinguishing between precise legal terms and words with fuzzy meanings. The resulting thesaurus can thus be seen as an important step to overcome the "untidiness" of natural language and to represent automatically the expert knowledge of a lawyer about legal terminology. Therefore, KONTERM can be used as a tool for the semi-automatic creation of a legal knowledge base.

1 Introduction

Legal expert systems have to face a very difficult task because the main prerequisite for their implementation - the formalism of law - has not been fulfilled yet. Formalism has always fascinated some legal scholars and legal informatics can be described as the last step to this aim.[1][15] In law there exists structured (logic, systems) and fuzzy knowledge (*grey areas*, poorly defined terms, natural language use, common-sense reasoning etc.). Whereas most approaches deal with the legal syllogism - the structured knowledge - our aim is a contribution to the problem of legal terms which are quite reasonable structured in legal theory however conventional databases and expert systems cannot represent this expert knowledge. Our aim is to develop a knowledge base on legal terms (structure of legal terms) which can be used for intelligent information retrieval and automatic indexing.

The importance of language for the law and the jurisprudence is without question.[9] Lawyers work with words, sentences, and texts. The language is a central subject of the lawyer's work. The particularity of the legal language is the use of a special vocabulary. The presence of a particular term has specific connotations. The result is that legal language is more precise than natural language. At best legal texts are clear, simple, and pithy. Legal thinking is based on the vocabulary of legal terms which are used to express a definite concept. Lawyers have formed with the theoretical methods of abstraction and logic thinking notions of human beings, objects, and processes.

Law has not been spared by the so called *information crisis*.[19] The then proposed information retrieval (IR)-systems supplied some help but did not solve the language problem. The property of legal vocabulary cannot be understood by the IR-system. As lawyers take the

use of legal language for granted in the process of working with legal texts and searching in legal databases the retrieval results are not so satisfactory. The IR-system cannot distinguish between legal terms and words with fuzzy meaning if the spelling is the same and thus produces much noise. The lawyer is not always aware of this deficiency of the database and also not able to overcome this problem by special research techniques.

2 Related work

Various approaches have been developed that deal with the formalisation of legal knowledge. Conflicting laws, statutes, and judicial decisions make the process of creating a rule-based expert system extremely difficult.[3][4][16] Open texture remains an unsolved problem for such systems although promising tools based on analogical and default reasoning have been proposed.[14]

Conceptual information retrieval systems use knowledge representation techniques to map the semantics of legal concepts.[5][7] Both graphical (e.g. conceptual graphs [8]) and structured (e.g. concept frames [6], objects [12]) schemes are applied. A recent extension is the connectionist approach which increases the adaptive behaviour. [11]

At the moment sophisticated techniques from natural language understanding are used only for small so-called question-answering systems.[10] A promising approach could be the application of simplified text analysis (information filtering). [2]

All the prototypes mentioned so far suffer from the great shortcoming that with respect to the high development costs the size of the information systems is so restricted that a commercial use is not feasible.

For the huge legal databases statistical methods appear to be the most promising instrument for automatic creation of selective legal thesauri. Most of the current work is based on the vector space model introduced by Salton/McGill [17]. Important extensions are the generalized vector space model [22], extended Boolean logic [18], and rough set approximation [20].

3 Basic concepts

KONTERM is a prototype of an intelligent IR-system which is designed to overcome this problem. The main hypothesis in the design of our prototype KONTERM is the assumption that a legal thesaurus can be used as a tool for automatic indexing if the applied descriptors are selective. If an intellectually created thesaurus is matched against document descriptors by means of Boolean retrieval logic the results are often noisy and unsatisfactory according to that missing selectivity. Therefore, our attempt to put aside this deficit was to check each thesaurus entry with regard to its selectivity to get a help for deciding whether the concerned term can be used as a precise descriptor. We also wanted to capture all distinct dimensions of meanings for a specific descriptor especially those *hidden* word senses which are not noticed in the process of intellectual indexing.

As legal terms are usually related to a special context KONTERM was designed to represent the meaning of a specific legal term by means of a vector. It is not calculated for the whole document but only for the surrounding context (i.e. 50 words before and after the relevant descriptor). Only significant nouns have been taken into account because especially for legal texts these terms contribute the major part to the meaning of the concerned context. The

similarity between two objects (text segments) is computed as a function of the number of properties that are assigned to both of them. According to the resulting similarity values the different occurrences of the descriptors are separated into various clusters. The surrounding text segments of the elements of a cluster form the basis for the process of cluster description. The ten most frequent terms occurring in these contexts are used for this purpose.

4 Implementation

The concepts presented in the preceding chapters were realised in a strictly modular and sequential way. We developed a process model which consists of eight different steps of analysis (see Figure 1).

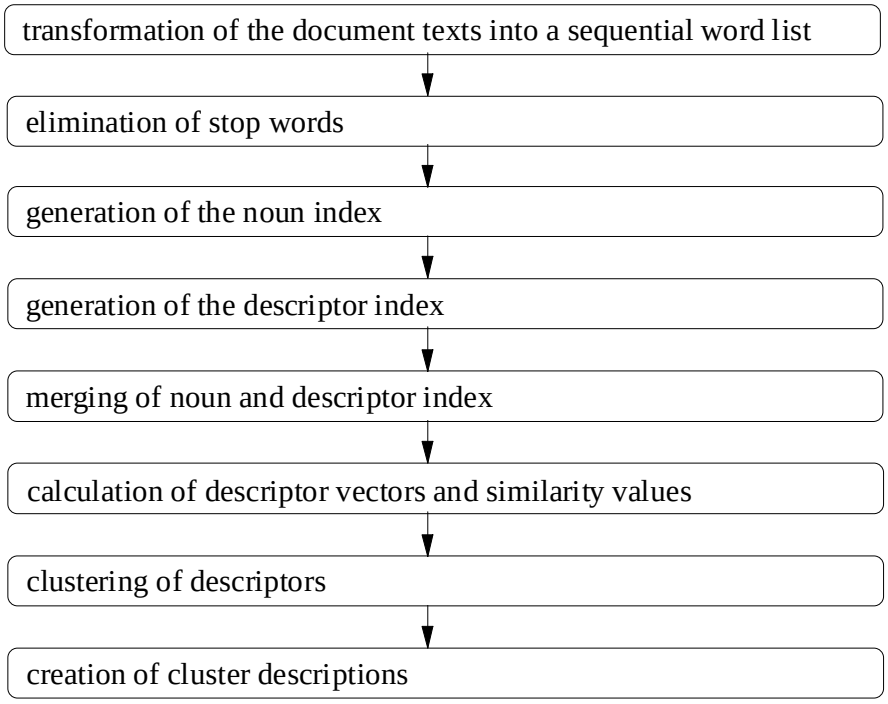

Figure 1: Process model

Each step in this model executes a well defined task and produces accurately specified output files which are in their turn the input for the following step of analysis. The relationships and dependencies between the different data files are shown in Figure 2. The dotted arrow between the word index and the stop word list signifies an intellectual transformation by means of a text editor whereas all other arrows represent automatic generations of the data files. So according to the principle of modular programming there exist modules with exactly determined interfaces between them which guarantees a consistent data exchange over all steps of the analysis process.

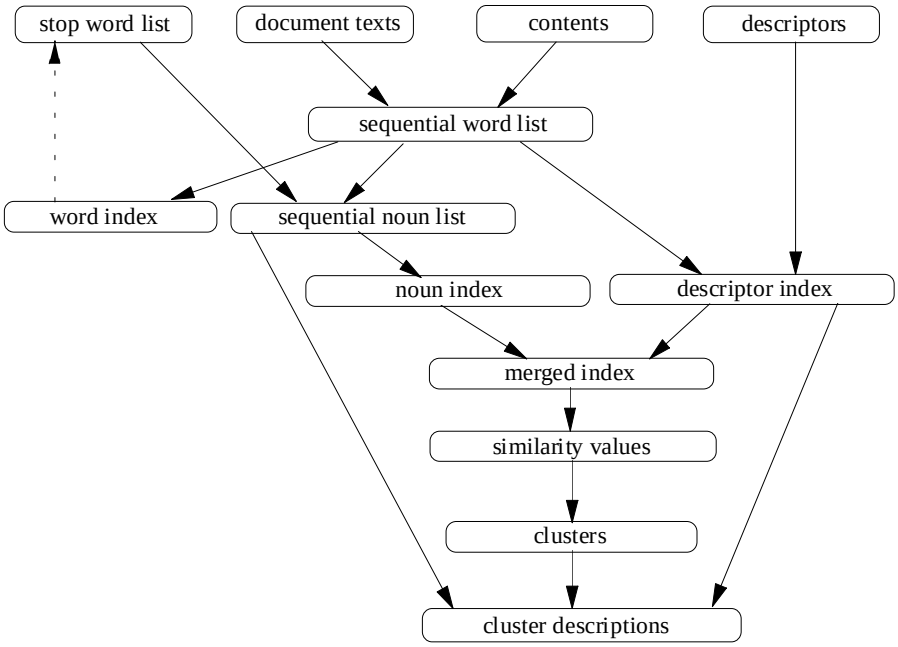

Figure 2: Relationships between data files

4.1 Transformation of the document texts into a sequential word list

The starting-point for the following steps of analysis represents a sequential word list which is produced based on the original document texts (ASCII format). For the generation the following rules are applied:

- *R1) A valid word consists only of alphabetic characters.* This heuristic turned out to be a good approximation especially for legal document texts. The only apparent point of weakness was the separation of the components of a compound word into several single words but for our aims this disadvantage was of no great importance.
- *R2) A valid word consists of at least two characters.* We introduced a lower bound for the word length because we wanted to eliminate all the meaningless abbreviations which consist only of one character or of several characters where each of them is separated by a dot. On the other hand the threshold was not set to a higher value in order to capture all significant abbreviations that could have an important influence for the meaning of context.
- *R3) The German special characters (ä, ö, ü, ß) are transformed to their international equivalent (ae, oe, ue, ss).* This conversion was necessary to standardise the different styles of writing, that is, there were documents which contained the original special characters and others with their international transcription. Additionally, the transformation was very helpful for the subsequent comparison of different inflexions of one word in the process of generating the noun index.
- *R4) All characters are converted to upper case letters.* The exclusive use of upper case letters was also inevitable in spite of the loss of information because there existed document texts which were printed only in upper case alphabetic characters. Furthermore this unification made the distinction between the normal and the substantivial use of a word as well as the use as the first word of a sentence (the latter two are written with a capital letter) superfluous.

4.2 Elimination of stop words

According to our basic assumption that nouns concentrate the meaning of a text segment all other word categories are eliminated. For that purpose a list of stop words has to be produced. To simplify the generation of that stop word list there exists the possibility to create a word index based on the sequential word list.

Within this word index each entry is supplemented with the number of its occurrences offered as support to decide if the concerned noun possesses the capability to influence the meaning of a specific context or if it is used only as an expletive. The process of eliminating the stop words takes the sequential word list and the stop word list as input and erases all occurrences of stop words in the former so that the result represents a reduced sequential list containing only the significant nouns.

4.3 Generation of the noun index

Based on the sequential noun list an index is produced where each entry is assigned with a list of all postings, that is, for each occurrence the document number and the position in the document. In order to test the equivalence of two words in the sequential word list we did not apply an exact string match but on the contrary a special module was developed which covers the following two cases:

- *Spelling errors*: Of course we could not take into account all possibilities for the appearance of spelling errors. Instead of that we treated the most frequent type of spelling error, namely the erroneous insertion of one wrong character. This kind of spelling correction covers also the vowel-gradation in the declension of German nouns. The only exception that we introduced was that we did not take into account a false insertion at the first position of a noun because there are some distinct nouns which differ from each other only in the first letter. Beyond that a wrong insertion of a letter in the first place is very unlikely and therefore can be neglected.
- *Word endings*: These are important phenomena for an inflective language like German. The distinct final parts of two analysed nouns are compared with a list of legal inflexions to determine their equality or diversity.

In the extensive testing of the comparison module we obtained a very high precision. Almost all different word shapes were captured whereas there was no incorrect assignment. We compared our method with other error correction algorithms which use e.g. thresholds for absolute or relative measures of divergence. In all cases we found our approach superior mainly because the used documents were already cleaned up to a high degree from spelling errors.

4.4 Generation of the descriptor index

In addition to descriptors which consist only of a single word the following two special cases are considered:

- *Compound descriptors*: For the indexation of legal documents descriptors which consist of several words are of particular importance. The compound term is inserted in the descriptor file in that the individual parts are simply separated by blanks. At first the entries

in the word list are compared with the first part of the descriptor term. If an equality is detected the second descriptor part is compared with the following word and so on until the whole compound descriptor is included in the comparative operation.

- *Synonyms*: In the case that distinct declensions of one descriptor can not be captured by the comparison module (irregular declensions) or that other terms with identical meaning for a descriptor exist, these synonyms can easily appended to the original descriptor in the descriptor file. The various synonyms are separated by means of equals signs. During the searching process each word in the sequential word list is compared with all synonyms to test the identity with one of them.

The descriptors are searched in the sequential word list, the resulting postings (document number and position) are assigned to the concerned descriptors to form the descriptor index. For the comparison of the words in the sequential word list with the descriptors the comparison module is applied again.

4.5 Merging of noun and descriptor index

The noun and the descriptor index are combined to build a merged index for each descriptor. For that purpose each noun posting is compared with each descriptor posting whether the noun is part of the context of the concerned descriptor. If the document numbers are the same and the relative position of the noun with respect to the descriptor position lies in the range of ± 50 words then the number of the descriptor posting is added to the new posting list of the noun for the descriptor in question (see Figure 3).

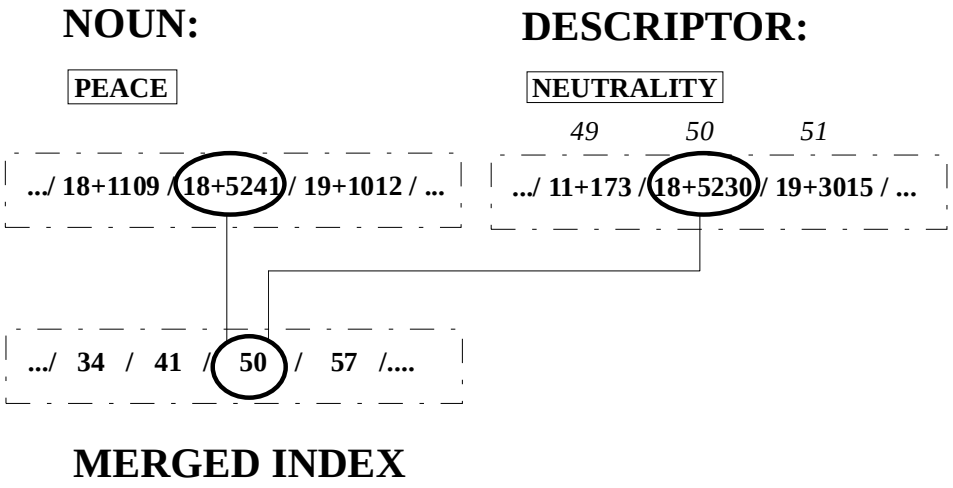

Figure 3: Example of merging noun and descriptor index

These merged indices are the basis for the following calculations by supplying the required information about the composition of the different contexts of a descriptor, that is, those nouns which determine the meaning of that specific use of the descriptor term.

4.6 Calculation of descriptor vectors and similarity values

According to the vector space model of information retrieval [17] for each descriptor occurrence *i* a vector is calculated which is supposed to capture its meaning as function of the presence or absence of certain nouns in its context, referred to as its properties:

$$D_i = (TERM_{i1}, TERM_{i2}, TERM_{i3}, \ldots, TERM_{it})$$

We used binary indexing so that $TERM_{ik} = 1$ if the noun k is present in the context of that descriptor occurrence and $TERM_{ik} = 0$ otherwise.[21] By means of the calculated descriptor vectors the similarity between two different occurrences of one descriptor can be interpreted as function of the number of properties which they have in common. This signifies that the similarity is directly proportional to the number of equal values of $TERM_{ik}$ for the components of the two descriptor vectors. For our analysis we used the symmetric similarity *coefficient of Dice* [17] because we attained the most reasonable results in comparison with other similarity measures (e.g. *coefficient of Jaccard*, *cosine coefficient* or *overlap measure*):

$$K(D_1, D_2) = \frac{2 \sum_{k=1}^{t} (TERM_{ik} \cdot TERM_{jk})}{\sum_{k=1}^{t} TERM_{ik} + \sum_{k=1}^{t} TERM_{jk}}$$

The value range of the *coefficient of Dice* remains within the interval *[0,1]*. If the coefficient equals *0* then the two vectors are completely dissimilar. On the other hand if it equals *1* the two vectors are identical. Therefore, the coefficient value represents exactly the percentage of nouns which are present in both descriptor contexts.

4.7 Clustering of descriptors

On the basis of the calculated similarity values between the different descriptor occurrences we tried to group them according to varying meanings so that we could capture distinct dimensions of word senses. The process of building groups of descriptor occurrences was achieved by means of a quick partition algorithm which created non-hierarchical disjunctive clusters.[13]

The applied method tests each pair of descriptor occurrences if its similarity value is greater than a pre-defined threshold. In this case if one of the two participants is already a member of a given cluster then the other is inserted to the already existing cluster elements. If both of the two descriptor occurrences are not included in any of the present clusters then a new cluster is created containing only these two entries. In a second run the resulting clusters are tested against each other for the occurrence of duplicate entries. If such a pair is detected the two clusters in question are joined into one. Finally, all descriptor occurrences which do not participate in any of the resulting clusters are inserted as additional clusters consisting of those single elements. In spite of that simple clustering algorithm we received very reasonable results. A lower limit of 20 for the coefficient of Dice yielded the best performance.

4.8 Creation of cluster descriptions

In the last step of our analysis process we supplemented the achieved clusters with descriptions to get a meaningful output which constitutes a representative description of the dimension of word sense for each cluster. For each cluster element we retrieved the associated descriptor posting from the corresponding descriptor index. These postings were searched in the sequential noun list. This resulted in the appropriate context for each descriptor posting.

All these document segments were the basis for an additional indexation process leading to an index of all nouns present in the contexts of the elements of a specific cluster. The ten most frequent terms were selected as output for the cluster description and ranked according to their number of occurrences.

5 Evaluation

5.1 Test database *Neutrality*

As a pre-test for the validity of our approach we used a small test database consisting of 56 text segments of documents from the European Community law database CELEX. The basis for our selection was a search in the database for the term *Neutralität* (neutrality). We applied our analysis not on whole documents but only on already prepared contexts for one descriptor in order to check in particular the last three steps of our implemented process model (i.e. calculation of descriptor vectors and similarity values, clustering, and creation of cluster descriptions). Whereas for an international lawyer *neutrality* seems to be used always in the sense of public international law, it was quite interesting to discover the other different meanings of this legal term which were detected by our prototype:

- neutrality of the public service (political aspects, sexual neutrality)
- neutrality of internal taxation (fiscal neutrality of the value-added tax system etc.)
- neutrality in the calculation and application of monetary compensatory amounts
- neutrality of the co-responsibility levy
- neutrality of the customs valuation system
- fiscal neutrality of refunds to exporters
- neutrality of the STABEX system
- chemical neutrality
- neutrality of competition

The same result can only be achieved by intellectual separation of the context windows which would be very time consuming as we noticed within the process of verification. Automatic indexing is also more precise because in practice indexers very often overlook some of the less common meanings.

5.2 Database *Austrian Treaties 1988 - 1992*

The database consists of 77 complete documents of Austrian public international law documents (mainly treaties) from 1988 to 1992. A representative list of 89 test descriptors has been created intellectually, both compound descriptors and synonyms were included. As a result of the application of our prototype system KONTERM, very useful output was produced. The following examples will show that all various meanings of a specific descriptor and also fuzzy uses of a term have been described automatically by our system according to our initial requirement.

For reasons of convenience the following cluster descriptions and the contexts have been translated into English. Therefore, the resulting inaccuracies (varying length of text segments, different terminology etc.) represent only the consequence of this translation procedure.

a) Legal term *Gleichberechtigung* (equality, equal rights):
13 text segments

Cluster 1: equality between States
7 text segments

Description: Moon, respect, principles, peoples, treaties, justice, public international law, co-operation, human rights, relations

2 representative text segments:

Charter of the United Nations	Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and other Celestial Bodies
WE THE PEOPLES OF THE UNITED NATIONS DETERMINED to save succeeding generations from the scourge of war, which twice on our lifetime has brought untold sorrow to mankind, and to reaffirm faith in fundamental human rights, in the dignity and worth of the human person, in the **equal rights** of men and women and of nations large and small, and to establish conditions under which justice and respect for the obligations arising from treaties and other sources of international law can be maintained, and to promote social progress and better standards of life in larger freedom, AND FOR	the interests of all countries, irrespective of their degree of economic or scientific development, and shall be the province of all mankind. Outer space, including the moon and other celestial bodies, shall be free for exploration and use by all States without discrimination of any kind, on a basis of **equality** and in accordance with international law, and there shall be free access to all areas of celestial bodies. There shall be freedom of scientific investigation in outer space, including the moon and other celestial bodies, and States shall facilitate and encourage international co-operation in such investigation. ARTICLE II Outer space, including

Cluster 2: equal rights of men and women
2 text segments

Description: Rights, covenant, enjoyment, notice, elimination, forms, discrimination, law on equal treatment, man, woman

Representative text segment:

International Covenant on Economic, Social and Cultural Rights
social origin, property, birth or other status. 3. Developing countries, with due regard to human rights and their national economy, may determine to what extent they would guarantee the economic rights recognised in the present Covenant to non-nationals. Article 3 The States Parties to the present Covenant undertake to ensure the **equal right** of men and women to the enjoyment of all economic, social and cultural rights set forth in the present Covenant. Article 4 The States Parties to the present Covenant recognise that, in the enjoyment of those rights provided by the State in conformity with the present Covenant, the State may subject

Cluster 3: reservations of former socialist States
3 text segments

Description: Reservation, mission, sending State, government, principle, view, dissent, agreement, receiving State, rules

Cluster 4: equality of the judge *ad hoc*
1 text segment

Description: Judges, court, president, allowance, one party only, for the purpose of, doubt, statute, conditions, decision

Text segment:

Statute of the International Court of Justice
one party only. Any doubt upon this point shall be settled by the decision of the Court. 6. Judges chosen as laid down in paragraphs 2,3 and 4 of this Article shall fulfil the conditions required by Articles 2, 17 (paragraph 2), 20, and 24 of the present Statute. They shall take part in the decision on term of complete **equality** with their colleagues. Article 32. 1. Each member of the Court shall receive an annual salary. 2. The President shall receive a special annual allowance. 3. The Vice-President shall receive a special allowance for every day on which he acts as President. 4. The judges chosen under Article 31, other than members of the Court,

So the four clusters as can be seen represent properly the different connotations of the term *Gleichberechtigung* (equality, equal rights).

b) Legal term *Arbeit* (work, carrying out, operating, activity, employment):
64 text segments

Cluster 1: fuzzy meaning
47 text segments

Description: Rights, contracting parties, protocol, protection, development, committee, implementation, service, security council, sessions

Representative text segment:

Agreement Establishing the European Bank for Reconstruction and Development
Board of Directors shall be responsible for the direction of the general operations of the Bank and, for this purpose, shall, in addition to the powers assigned to it expressly by this Agreement, exercise all the powers delegated to it by the Board of Governors, and in particular: (i) prepare the **work** of the Board of Governors, (ii) in conformity with the general directions of the Board of Governors, establish policies and take decisions concerning loans, guarantees, investments in equity capital, borrowing by the Bank, the furnishing of technical

Cluster 2: enterprises operating competitively 3 text segments

Description: Enterprises, assistance, transition, private ownership, control, participation, market oriented economy, total amount, country, competitively

Agreement Establishing the European Bank for Reconstruction and Development such a period: (a) the Bank shall provide to such a country, and to enterprises in its territory, upon their request, technical assistance and other types of assistance directed to finance its private sector, to facilitate the transition of state-owned enterprises to private ownership and control, and to help enterprises **operating** competitively and moving to participation in the market oriented economy, subject to the proportion set forth in paragraph 3 of Article 11 of this Agreement; (b) the total amount of any assistance thus provided shall not exceed the total amount of cash disbursed and promissory notes issued by that country for	Agreement Establishing the European Bank for Reconstruction and Development determining their use; and (v) by making or participating in loans and providing technical assistance for the reconstruction or development of infrastructure, including environmental programmes, necessary for private sector development and the transition to a market-oriented economy. For the purposes of this paragraph, a state-owned enterprise shall not be regarded as **operating** competitively unless it operates autonomously in a competitive market environment and unless it is subject to bankruptcy laws. 2. (i) The Board of Directors shall review at least annually the Bank's operations and lending strategy in each recipient country to ensure that the purpose and the functions of the Bank, as set out in Articles 1 and	Agreement Establishing the European Bank for Reconstruction and Development the Articles of Agreement of the Bank. During that period, the Soviet Union wishes that the Bank will provide technical assistance and other types of assistance directed to finance its private sector, to facilitate the transition of state-owned enterprises to private sector ownership and control and to help enterprises **operating** competitively and moving to participation in the market-oriented economy, subject to the proportion set forth in paragraph 3 of Article 11 of this Agreement. The total amount of any assistance thus provided by the Bank would not exceed the total amount of the cash disbursed and the promissory notes issued

Cluster 3: termination of activities 1 text segment

Description: Termination, operations, liability, claims, Board of Governors, opportunity, action, affirmative vote, Governors, members

Agreement Establishing the European Bank for Reconstruction and Development opportunity for further consideration and action by the Board of Governors. Article 41 Termination of Operations The Bank may terminate its operations by the affirmative vote of not less than two-thirds of the Governors, representing not less than three-fourths of the total voting power of the members. Upon such termination of operations the Bank shall forthwith cease all **activities**, except those incident to the orderly realisation, conservation and preservation of its assets and settlement of its obligations. Article 42 Liability of Members and Payment of Claims 1. In the event of termination of the operations of the Bank, the liability of all members for uncalled subscriptions to the capital stock of

Cluster 4: taxation of employment

6 text segments

Description: Remuneration, services, alienator, profession, pensions, property, directors' fees, similar payments, political subdivision

Representative text segment:

Agreement between the Government of the Republic of Austria and the Government of Malaysia for the Avoidance of Double Taxation and the Prevention of Fiscal Evasion with Respect to Taxes on Income

those mentioned in paragraphs 1 and 2 of this Article, shall be taxable only in the Contracting State of which the alienator is a resident.
Article 14
Personal Services
1. Subject to the provisions of Articles 15, 16, 17 and 18 remuneration derived by an individual who is a resident of a Contracting State in respect of an employment or a profession shall be taxable only in that State unless the **employment** or profession is so exercised, such income as is derived therefrom may be taxed in that other State.
2. Notwithstanding the provisions of paragraph 1 remuneration derived by an individual who is a resident of a Contracting State in respect of such employment or profession exercised in the Contracting State shall

Cluster 5: preparatory work as supplementary means of interpretation
1 text segment

Description: Meaning, parties, interpretation, application, treaty, means of interpretation, practice, agreement, relations, rules of international law

Text segment:

Vienna Convention on the Law of Treaties

(b) any subsequent practice in the application of the treaty which establishes the agreement of the parties regarding its interpretation;
(c) any relevant rules of international law applicable in the relations between the parties.
4. A special meaning shall be given to a term if it is established that the parties so intended.
Article 32
Supplementary means of interpretation
Recourse may be had to supplementary means of interpretation, including the preparatory **work** of the treaty and the circumstances of its conclusion, in order to confirm the meaning resulting from the application of article 31, or to determine the meaning when the interpretation according to article 31:
(a) leaves the meaning ambiguous or obscure; or
(b) leads to a result which is manifestly absurd or

Each of the remaining seven clusters consists of one single element, so we give in the following only a brief description of their meaning:

- *Cluster 6:* ERASMUS Program Agreement - costs of preparatory work
- *Cluster 7:* IAEA Convention on Assistance - co-operation on assistance
- *Cluster 8:* IAEA Convention on Assistance - co-operation on compensation of claims
- *Cluster 9:* UN Convention on Contracts for the International Sale of Goods: non application on work orders
- *Cluster 10:* Buthan Agreement on Technical Co-operation: working permits
- *Cluster 11:* Treaty on the Establishment of a Culture Institute - application of the labour and social regulations
- *Cluster 12:* EDI Agreement: current activities in case of notice

The main cluster represents the fuzzy meaning of the term *Arbeit* which should therefore not be used as a descriptor. The other 11 clusters reflect the various specific connotations of the term. As could be demonstrated *Arbeit* is an example of a very general descriptor but it possesses also meanings that are to a high degree precise and selective and which can be detected by means of our prototype KONTERM.

6 Conclusion and future work

The clusters as well as the cluster descriptions can be considered as very helpful for the development of a legal thesaurus. KONTERM can exactly distinguish between the usage of a descriptor as a specific legal term and its general meaning. We also captured all distinct connotations for a specific descriptor. This automatic indexing of legal terms by means of this selective thesaurus is much more precise and efficient than traditional indexing techniques. In particular those *hidden* word senses are detected which are not noticed in the process of intellectual indexing.

The resulting selective legal thesaurus together with the corresponding cluster descriptions can be used to automatically supplement the individual documents with representative descriptions. These condensed representations could be linked to the concerned documents and would constitute an important assistance for the user of the full text information retrieval system.

Major goals of future research are to improve the representation of legal knowledge and the analysis of new documents. Based on the output of our knowledge acquisition tool we will investigate various other approaches in order to achieve a conceptual knowledge base even for huge data material.

The results of our extensive testing cycles were quite motivating. KONTERM satisfied to a high degree our initial expectations. It represents an efficient assistance to overcome the problem of automatic formalisation of legal language and a valuable representation tool for well structured expert knowledge about legal terminology.

Acknowledgements

Our work is funded by the Jubiläumsfonds of the Oesterreichische Nationalbank - research project no. 4225: *Expert System for Public International Law*.

References

[1] Allen L.E., Towards a Normalized Language to Clarify the Structure of Legal Discourse, in: Martino A.A. (ed.), "Deontic Logic, Computational Linguistics and Legal Information Systems", Amsterdam, North Holland, 1982, p. 349-405.

[2] Appelt D.E. et al., FASTUS: A Finite-state Processor for Information Extraction from Real-world Text, in: Bajcsy R. (ed.) "13th Int. Joint Conf. on Artificial Intelligence", San Mateo CA, Kaufmann, 1993, p. 1172-1178.

[3] Baaz M. et al., The Application of Kripke-Type Structures to Regional Development Programs, in: Marík V., Lazansky J., and Wagner R.R. (eds.) "4th Int. Conf. on Database and Expert Systems Applications", Berlin, Springer, 1993, p. 523-528.

[4] Bench-Capon T.J.M. (ed.), Knowledge Based Systems and Legal Applications, London, Academic Press, 1991.

[5] Bing, J., Designing Text Retrieval Systems for "Conceptual Searching", "1st Int. Conf. on Artificial Intelligence & Law", Baltimore, ACM Press, 1987, p. 43-51.

[6] Croft W.B., Thompson R.H., I^3R: A New Approach to the Design of Document Retrieval Systems, JASIS, Vol. 38, 1987, p. 389-404.

[7] Cross G.R., DeBessonet C.G., Representation of Legal Knowledge for Conceptual Retrieval, Information Processing & Management, Vol. 21, 1985, p. 35-44.

[8] Dick, J.P., Representation of Legal Text for Conceptual Retrieval, "3th Int. Conf. on Artificial Intelligence & Law", Baltimore, ACM Press, 1991, p. 244-252.

[9] Hatz H., Rechtssprache und juristischer Begriff, Stuttgart, Kohlhammer, 1963.

[10] Jacobs P.S., Rau L.S., SCISOR: Extracting Information from On-line-News, CACM, Vol. 33, No. 11, 1990, p. 88-97.

[11] Jaquemin C., A Coincidence Detection Network for Spatio-Temporal Coding: Application to Nominal Composition, in: Bajcsy R. (ed.) "13th Int. Joint Conf. on Artificial Intelligence", San Mateo CA, Kaufmann, 1993, p. 1346-1351.

[12] Mital V., Stylianou A., and Johnson L., Conceptual Information Retrieval in Litigation Support Systems, "3th Int. Conf. on Artificial Intelligence & Law", Baltimore, ACM Press, 1991, p. 235-243.

[13] Panyr J., Vektorraum-Modell und Clusteranalyse in Information-Retrieval-Systemen, Nachr. Dok., Vol. 38, 1987, p. 13-20.

[14] Quirchmayr G., Open Texture and Knowledge Representation, in: Schmitz K.-D. (ed.) "3rd Int. Congress on Terminology and Knowledge Engineering", Frankfurt/M., Indeks, 1993, p. 87-91.

[15] Reisinger L., Rechtsinformatik, Berlin, 1977.

[16] Rissland E.L., Skalak D.B., and Friedman M.T., Case Retrieval through Multiple Indexing and Heuristic Search, in: Bajcsy R. (ed.) "13th Int. Joint Conf. on Artificial Intelligence", San Mateo CA, Kaufmann, 1993, p. 902-908.

[17] Salton G., McGill M.J., Introduction to Modern Information Retrieval, New York, McGraw-Hill, 1983.

[18] Salton G., A Simple Blueprint for Automatic Boolean Query Processing, Information Processing & Management, Vol. 24, 1988, p. 269-280.

[19] Simitis S., Informationskrise des Rechts und Datenverarbeitung, Karlsruhe, Müller, 1970.

[20] Srinivasan P., Intelligent Information Retrieval using Rough Set Approximations, Information Processing & Management, Vol. 25, 1989, p. 347-361.

[21] Turtle H.R., Croft W.B., A Comparison of Text Retrieval Models, Computer Journal, Vol. 35, 1992, p. 279-290.

[22] Wong S.K.M. et al., On Modeling of Information Retrieval Concepts in Vector Spaces, ACM Trans. on Database Systems, Vol. 12, 1987, 299-321.