Artificial Intelligence and Law **7:** 185–209, 1999.

Exploratory analysis of concept and document spaces with connectionist networks ⋆

DIETER MERKL
Institute of Software Technology, Vienna University of Technology, Resselgasse 3/188, A-1040 Vienna, Austria

ERICH SCHWEIGHOFER
Institute of Public International Law, University of Vienna, Universitätsstraße 2, A-1090 Vienna, Austria

WERNER WINIWARTER
Department of Information Science, Kyoto University, Sakyo, Kyoto, 606-01 Japan

Abstract. Exploratory analysis is an area of increasing interest in the computational linguistics arena. Pragmatically speaking, exploratory analysis may be paraphrased as natural language processing by means of analyzing large corpora of text. Concerning the analysis, appropriate means are statistics, on the one hand, and artificial neural networks, on the other hand. As a challenging application area for exploratory analysis of text corpora we may certainly identify text databases, be it information retrieval or information filtering systems. With this paper we present recent findings of exploratory analysis based on both statistical and neural models applied to legal text corpora. Concerning the artificial neural networks, we rely on a model adhering to the unsupervised learning paradigm. This choice appears naturally when taking into account the specific properties of large text corpora where one is faced with the fact that input-output-mappings as required by supervised learning models cannot be provided beforehand to a satisfying extent. This is due to the fact of the highly changing contents of text archives. In a nutshell, artificial neural networks count for their highly robust behavior regarding the parameters for model optimization. In particular, we found statistical classification techniques much more susceptible to minor parameter variations than unsupervised artificial neural networks. In this paper we describe two different lines of research in exploratory analysis. First, we use the classification methods for concept analysis. The general goal is to uncover different meanings of one and the same natural language concept. A task that, obviously, is of specific importance during the creation of thesauri. As a convenient environment to present the results we selected the legal term of "neutrality", which is a perfect representative of a concept having a number of highly divergent meanings. Second, we describe the classification methods in the setting of document classification. The ultimate goal in such an application is to uncover semantic similarities of various text documents in order to increase the efficiency of an information retrieval system. In this sense, document classification has its fixed position in information retrieval research from the very beginning. Nowadays renewed massive interest in document classification may be witnessed due to the appearance of large-scale digital libraries.

Key words: artificial neural networks, exploratory data analysis, legal information retrieval, natural language processing, unsupervised learning, vector space model

1. Introduction

The information crisis in law (Simitis, 1970) was the impetus for the development of legal information retrieval systems. As a result of a first huge effort a number of information retrieval systems have been developed with sufficient coverage concerning the underlying text corpora. Lawyers, however, need much more than just a documentation of the various legal acts of the relevant jurisdiction. The material, rather, has to be organized in a systematic manner in the form of a legal commentary. In this regard, we have to confess that the level of quality for useful systems is set rather high with respect to the more than 2500 years of intellectual experience. This experience, obviously, represents the baseline against which improvements have to be measured. The major question for such systems is to find an efficient way to formalize legal knowledge. A number of different approaches can be distinguished. Some of the more influential ones are outlined below.

The first solution was to utilize various document types and fields representing semantic knowledge as a tractable means to represent deep structure in legal information retrieval systems (Schweighofer, 1995). Two major problems remain unsolved, namely first, the users are not experienced enough to deal with these difficult but efficient search algorithms, and second, the documents have to be indexed manually thus making the development highly time-consuming. Similar problems occur when adhering to a knowledge-based approach (Bing, 1987; Cross and Bessonet, 1985) towards legal information representation. Hence, a time-consuming manual construction of the knowledge-base is independent of the actual mechanism for encoding the semantics of legal concepts, thus irrespective of a particular knowledge representation technique be it semantic networks (Paice, 1991), conceptual graphs (Dick, 1991), concept frames (Hafner, 1981), diagnostic expert systems (Merkl *et al.*, 1992), object-oriented programming (Mital *et al.*, 1991) or case-based reasoning (Ashley, 1990). The other side of the coin, obviously, is marked by improved capabilities of the overall system with respect to retrieval efficiency.

Neural networks found some attention for encapsulation of legal knowledge. This might be due to the fact of only limited success of knowledge-based approaches. Two main streams of research may be observed. First, neural networks are trained to represent vague concepts according to some predefined input-output-mapping (Bench-Capon, 1993; Philipps, 1989) or structured schematic interpretations of case descriptions as suggested in (Groendijk, 1992; Groendijk and Oskamp, 1993; van Opdorp *et al.*, 1991). Second, neural networks are used to perform a spreading-activation during retrieval as another paradigm to describe the relation between terms, on the one hand, and documents and queries, on the other hand (Belew, 1987; Rose and Belew, 1989; Rose, 1994). We have to note, however, a severe disadvantage of such spreading-activation models when they are

* The authors are affiliated to the Research Center for Computers and Law at the Institute of Public International Law.

compared with inference networks (Turtle and Croft, 1991). In particular, while both approaches yield similar results the inference network is considerably less demanding as far as the necessary computing power is concerned. The main advantage of learning in connectionist systems, on the other hand, has been realized in small experimental settings only.

Some compromise between exact knowledge representation and ease of use as well as ease of creation is represented by non-linear networks or hypertext to use a more common term (Greenleaf *et al.*, 1995). The self-explanation of the non-linear network may offer deep structure of knowledge representation also to inexperienced users. The manual construction of knowledge-bases and the resulting timely demands are addressed by automatic indexing of full-text documents. Convincing results in this line of research using inference networks (Turtle and Croft, 1991) are reported in (Turtle, 1995).

Legal thinking and logic programming are not that similar as early work suggested (Sergot *et al.*, 1986; Susskind, 1987). First-order logic is appropriate only for particular applications. The huge quantity of legal rules makes coding and maintenance a tricky issue. Deontic logic, on the other hand, is not yet in a state of maturity that would allow real-world applications.

A different stream of research in legal information retrieval is linked with the field of (legal) computational linguistics. Most approaches use the vector space model (Salton and McGill, 1983; Salton, 1989) in order to formalize documents (Mulder and van Noortwijk, 1994), citation vectors (Tapper, 1982) or document descriptions (Smith *et al.*, 1995). Only recently, some work relying on the utilization of the particularities of natural language found considerable attraction in the information retrieval area (Deerwester *et al.*, 1990; Maarek and Smadja, 1989). Concerning the specific requirements of legal applications, recent work relies on extracts and templates (Gelbart and Smith, 1993; Smith *et al.*, 1995), linguistic constructs (Konstantinou *et al.*, 1993), and legal concepts (Schweighofer *et al.*, 1995) as an appropriate form of abstraction of legal language. Lawyers have formed definite concepts of human beings, objects, and processes by use of methods of abstraction and logic thinking. Concepts can be formalized with context-sensitive terms or rules. Combined with statistics or neural networks for purposes of knowledge acquisition, a powerful tool for the semi-automatic description of documents or terminological analysis is available. Such a method of exploratory data analysis (Church and Mercer, 1993) is an effective and computationally tractable tool for summarizing and clustering legal documents.

In the spirit of this line of argumentation, the design considerations during the KONTERM project have been fixed (Schweighofer and Winiwarter, 1993a; Merkl *et al.*, 1994; Schweighofer *et al.*, 1995). The backbone of the project consists of a database of text patterns representing legal concepts. The text corpus is analyzed by means of linguistic, statistical, and connectionist methods in order to achieve, on the one hand, selective descriptors to be used in legal thesauri (Schweighofer and Winiwarter, 1993b) and, on the other hand, a classification schema for legal documents. One critical problem during this analysis is the detection of the connotations of each descriptor. The connotation analysis is based on the interpretation of the contexts of the individual legal terms. To calculate the similarity between descriptor occurrences we applied both statistical techniques as well as neural networks in order to uncover different word meanings. The results of connotation analysis are further used for document representation in that legal documents are represented by means of terms extracted from a particular document as well as the connotations of that very term. Finally, document clusters are formed with respect to the similarity between their respective representation.

The material presented in the remainder of this paper is organized as follows. In Sect. 2 we provide a brief overview of the field of artificial neural networks. Major emphasis is directed to models of unsupervised learning since we rely on unsupervised learning for concept and document analysis. Sect. 3 contains an in-depth exposition of the formalism of concept and document representation that acts as the basis for the case studies described in Sect. 4. Finally, we present some conclusions in Sect. 5.

2. Connectionist Networks

2.1. BASIC CONSIDERATIONS

Research in the area of artificial neural networks dates back to the early 1940's when Warren S. McCulloch and Walter Pitts described a conceptual model of electronic circuits performing computational tasks, inspired by a model of biological neurons (McCulloch and Pitts, 1943). In the late 1950's the first computational models appeared in literature. The whole research area, however, was hit back to almost non-existence by the publication of a single book (Minsky and Papert, 1969) which disclosed the deficiencies of the computational models. A renewed interest in artificial neural networks started in the 1980's, caused by a learning rule (Rumelhart *et al.*, 1986; Werbos, 1974; Werbos, 1994) which overcomes the limitations of the early models. This section contains a brief overview of the basic principles of artificial neural networks and a review of some unsupervised learning architectures. Good sources for introductory information about artificial neural networks are (Bishop, 1995; Jain and Mao, 1996; Kohonen, 1988; Pao, 1988; Ripley, 1996).

The basic elements of information processing in artificial neural networks are often termed neurons in analogy to the nervous system. However, due to their rather metaphorical similarity we prefer the terms computational neuron or unit for the sake of simplicity. Typically, an artificial neural network consists of a large number of interconnected units, each of which performs very simple operations. These operations may be characterized as follows. First, each unit receives input from a number of other units. These inputs are collected by using an input function. Second, the value of the input function representing the net-input to a unit at a given time is transformed by using a so-called activation function yielding the activation level or the state of this very unit. Third, the activation level of a unit is further transformed by means of an output function to obtain the output of a unit which in turn is input to a number of other units.

Data processing within artificial neural networks is performed by communication between various units via weighted connections. In other words, the output of one unit is propagated to a number of other connected units. The weights of the various connections represent the strength of the mutual connection between the two units concerned. Just to draw a biological analogy we might consider the weights as being the synaptic strength of the connection. To conclude, the information contained in an artificial neural network is stored in the connections between the units or more precisely in the weights assigned to these connections. Moreover, inherent in artificial neural networks is their highly parallel nature of data processing.

Artificial neural network models are specified by the network topology, the unit characteristics, and the learning rule. The network topology refers to the arrangement of units and their mutual connections. The unit characteristics consist of their input, activation, and output function. Finally, the learning rule determines the adaptation policy of the connection weights in order to improve the performance of the artificial neural network. In this sense the term adaptation refers to weight changes. However, some models have recently been suggested which enlarge the notion of adaptation to cover the network topology as well, see (Blackmore and Miikkulainen, 1995; Carpenter and Grossberg, 1988; Fritzke, 1994; Fritzke, 1995; Fritzke, 1996; Köhle and Merkl, 1996) to name but a few.

If we want to give a taxonomy of artificial neural network models we have in general a number of possibilities regarding, for example, the values of the input data, i.e. either binary or analog, or the arrangement of the connections, i.e. either one-directional or recurrent. In this paper, however, will give such a taxonomy in terms of the learning rule, i.e. either supervised or unsupervised. The basic distinction is made whether or not any externally supplied information concerning the correctness of the neurally computed output is provided during the learning phase of the artificial neural network. In other words, a learning rule which is based on the comparison between the actually computed output of the artificial neural network and a desired output is called supervised. Contrary to that, when no such desired output is specified and thus, no comparison takes place, one speaks of an unsupervised learning rule. More precisely, supervised learning is commonly performed by repeatedly presenting the input data to the artificial neural network, determining the network's output, and successively reducing the remaining deviation between the computed and the desired output. The learning process is terminated as soon as, for example, a tolerable remaining error is realized. Artificial neural network models adhering to the unsupervised learning paradigm do not get any additional information such as a desired output. Instead, they capture the regularities present in the input data. Pragmatically speaking, we might state that these models build their own representation of the input domain.

Due to the fact that our approach to reasoning in concept and document spaces relies on unsupervised neural networks we will give a somewhat deeper account on these models in this paper and omit an exposition of supervised neural networks. The interested reader will find an excellent treatment of mainstream supervised models in (Bishop, 1995).

2.2. UNSUPERVISED LEARNING IN ARTIFICIAL NEURAL NETWORKS

2.2.1. *Competitive learning*

Competitive learning (Rumelhart and Zipser, 1986), or *winner-takes-all* as it is termed quite often, may be regarded as the basis of almost all unsupervised learning strategies. In its basic form, a competitive learning network consists of k units with weight vectors m_i, $1 \leq i \leq k$, of equal dimension as the input data, $m_i \in \Re^n$. During learning, the unit c with its weight vector being the closest to the input vector x in terms of some distance metric, e.g. Euclidean distance, is selected. This very unit is further adapted in such a way that its weight vector resembles the input vector more closely. The unit containing this weight vector is dubbed the *winner* of the selection process. Obviously, the term *winner-takes-all* refers to the fact that only the winner is adapted whereas all other units remain unchanged. Such a learning rule may be established by gradually reducing the difference between weight and input vector. The actual amount of distance reduction is guided by means of a so-called learning-rate ϵ in the interval of [0, 1]. In such an environment the weight vectors tend to represent the mean of the input data matched onto that very unit. For the sake of clarity we provide the formulae for selection and adaptation in Equation 1 and Equation 2, respectively. In these equations we use a discrete time notation with t representing the time-stamp of the current learning iteration.

$$c : ||x(t) - m_c(t)|| \leq ||x(t) - m_i(t)||, \qquad 1 \leq i \leq k. \tag{1}$$

$$m_c(t+1) = m_c(t) + \epsilon \cdot [x(t) - m_c(t)]. \tag{2}$$

As a severe limitation of basic competitive learning consider the case where some units are never selected as the winner, due to the random initialization of their weight vectors. Consequently, their weight vectors are never changed. Such units may be referred to as *dead units* since they do not contribute to the learning process and thus, they do not contribute to input data representation.

This observation has led to a number of learning rules and network architectures that overcome this limitation by adapting not only the winner, see for instance the work reported in (Blackmore and Miikkulainen, 1995; Fritzke, 1994; Fritzke, 1995; Kangas *et al.*, 1990; Kohonen, 1982; Martinetz *et al.*, 1993; Sirosh and Miikkulainen, 1993; Sirosh and Miikkulainen, 1994; Wan and Fraser, 1994). Apart from the winner, adaptation is performed with units in some defined vicinity around the winner. This type of learning rules may be referred to as *soft competitive learning* rules. One of the most prominent representatives of this kind of artificial neural networks – the self-organizing map – will be presented below.

2.2.2. *Self-organizing maps*

The architecture of a self-organizing map as proposed by Teuvo Kohonen (Kohonen, 1982; Kohonen, 1989; Kohonen, 1990; Kohonen, 1995) consists of a layer of n input units and a grid of k output units each of which has assigned an n-dimensional weight vector m_i, $m_i = (\mu_1, \mu_2, \ldots, \mu_n)$. The task of the input units is to receive the various input patterns x, $x = (\xi_1, \xi_2, \ldots, \xi_n)$, representing real-world entities and to propagate them as they are onto the grid of output units. In parenthesis we should note at this point that the input patterns of the artificial neural network are the vectors representing the various legal concepts or legal documents. The description of such a vector representation for concepts and documents, however, is deferred to Sect. 3 in order to keep the exposition of the neural network at this stage as problem or application independent as possible.

Each of the output units in its turn computes exactly one output value which is proportional to the similarity between the current input vector and that unit's weight vector. This value is commonly referred to as the unit's activation or the unit's response to the presentation of an input. Usually, the Euclidean distance is used as the measure of similarity as shown in Equation 3. The model is, however, not restricted to the utilization of this particular metric, other similarity or dissimilarity measures should work equally well.

$$||x - m_i|| = \sqrt{\sum_{j=1}^{n}(\xi_j - \mu_j)^2}. \tag{3}$$

The adaptation of the weight vectors represents the crucial part of any unsupervised learning rule. This process may be described in three steps which are to be performed repeatedly. These three steps are henceforth collectively referred to as one learning iteration. First, one input vector at a time is randomly selected from the set of input vectors. Second, this input vector is mapped onto the grid of output units of the self-organizing map and the unit with the strongest response is determined. This very unit is further referred to as the winning unit, the winner in short. Notice that in case of Euclidean distance metric the unit with the smallest distance between input and weight vector is selected as the winner. Hence, the winner is the output unit representing the most similar internal representation of the input at hand. Third, the weight vector of the winner as well as weight vectors of units in topological neighborhood of the winner are adapted in such a way that these units will exhibit an even stronger response to the same input vector at future presentations. In less bulky terms, the third step refers to the reduction of distance between input and weight vectors of a subset of the output units and thus, to the improved correspondence between the description of an input and its internal representation. Such a distance reduction may easily be accomplished by a gradual reduction of the difference between corresponding vector components. This adaptation is further guided by a so-called learning-rate ϵ in the interval [0, 1] determining the amount of adaptation and a so-called neighborhood-rate $\psi_{c,i}$ determining the spatial range of adaptation.

In order to guarantee the convergence of the learning process, i.e. a stable arrangement of weight vectors, the learning-rate as well as the neighborhood-rate have to shrink in the course of time. In other words, the amount of adaptation of weight vectors decreases during the learning process with respect to a decreasing learning-rate. Furthermore, the amount of units that are subject to adaptation, i.e. the spatial range of adaptation, decreases as well during the learning process such that towards the end of learning only the winner is adapted and the weight vectors of neighboring units remain unchanged. Given these two restrictions it is obvious that the learning process will converge towards a stable arrangement of weight vector entries. Moreover, the self-organizing map will assign highly similar input data to neighboring output units thanks to the inclusion of a spatial dimension to the learning process. Hence, the self-organizing map represents a spatially smooth neural version of k-means clustering (Ripley, 1996).

With similar notation as defined in the paragraph on competitive learning above we may describe the rule for weight vector adaptation of unit i in the neighborhood of the winner c as given in Equation 4.

$$m_i(t+1) = m_i(t) + \epsilon(t) \cdot \psi_{c,i}(t) \cdot [x(t) - m_i(t)]. \tag{4}$$

For the sake of simplicity we omit the exact formulation of the neighborhood-rate $\psi_{c,i}$; for a detailed exposition of different realizations we suggest to consult (Kohonen, 1989; Kohonen, 1995; Ritter and Kohonen, 1989; Merkl, 1995b; Miikkulainen, 1991). A comparative study of the effects of different neighborhood-rates may be found in (Merkl, 1995c).

A simple graphical representation of a self-organizing map is provided in Fig. 1. In this figure the grid of output units consists of a square of 36 output units. One input vector x is mapped onto the grid of output units and the winning unit is selected. In the figure the winner is depicted as a black node. The weight vector of the winner, i.e. $m_c(t)$, is moved towards the current input vector. Since the input and the weight vector have equal dimension they may both be regarded as vectors of the same space and thus, both are depicted as belonging to the input space in the figure. As a consequence of the adaptation, the winner will produce a higher response with the same input vector at the next learning iteration, i.e. $t+1$, because the unit's weight vector, i.e. $m_c(t+1)$, is now nearer to the input vector x. Apart from the winner, adaptation is performed for neighboring units, too. Units that are subject to adaptation are depicted as shaded nodes in the figure. Moreover, the shading of the nodes corresponds to the degree of adaptation and thus, to the spatial

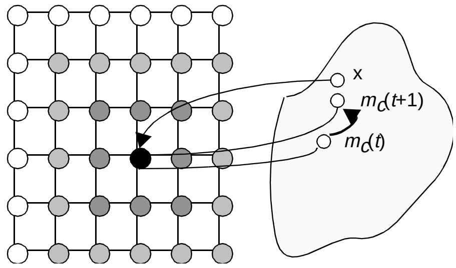


Figure 1. Self-organizing map.

range of weight vector adaptation. Generally, units in close vicinity to the winner are adapted more strongly and consequently, they are depicted with a darker shade.

For a more detailed description of the learning process as well as some variations consult (Kangas *et al.*, 1990; Kaski and Lagus, 1996; Kohonen, 1993; Kohonen, 1995; Merkl, 1995c; Merkl, 1995b).

The self-organizing map has already some tradition in information retrieval applications. Perhaps among the first to use self-organizing maps in this area we have to cite (Lin *et al.*, 1991). In this work the authors report on document clustering, where the documents are represented by means of 25 index terms taken from the various titles of scientific papers related to computer science. Hence, the vocabulary used in this study is fairly small. This line of research is continued, yet this time with full-text indexed documents that are consequently represented by means of a much larger and thus more realistic vocabulary as described in (Merkl, 1995a). Another stream of research is dedicated to the usage of self-organizing maps for document representation (Honkela *et al.*, 1995) based on the seminal work of (Ritter and Kohonen, 1989). Such a document representation is further used for clustering, again by means of the self-organizing maps (Lagus *et al.*, 1996).

3. Representation of Legal Text Corpora

3.1. BASIC CONSIDERATIONS OF COMPUTATIONAL LINGUISTICS

During the last years we witnessed a growing interest in computational linguistic technologies and their application to real-world systems (for a good survey see (Church and Rau, 1995)). Moreover, we observe a remarkable recent change of attitude in research on this domain away from the “toy problem syndrome” to the construction of scaleable real-world end-user applications. This paradigm shift resulted in the new label of *language engineering* (Cunningham *et al.*, 1996) and reflects the change of emphasis from the knowledge-based approach in computational linguistics (e.g. see (Allen, 1987)) to the increased use of corpus-based techniques (often also called *corpus-based linguistics*) (Armstrong, 1994; Charniak, 1993; Sharkey, 1992; Wermter, 1995). However, after the first failures of pure statistical or connectionist systems, many researchers settled down to the more realistic view of *exploratory data analysis* (Church and Mercer, 1993) which states that only the careful semi-automatic combination of empirical methods and intellectual evaluation can lead to the development of efficient and high-quality linguistic tools.

By following this line of argumentation we construct a linguistic model for legal concepts and documents by making use of connectionist and statistical techniques. Related work at the concept level is mainly focused on word sense disambiguation (Agirre and Rigau, 1996; Bruce and Wiebe, 1994; Karov and Edelman, 1996; Kozima and Furugori, 1993; Ng and Lee, 1996) and the automatic acquisition of thesauri (Grefenstette, 1994; Jing and Croft, 1994; Pereira *et al.*, 1993; Tokunaga *et al.*, 1995). At the document level research efforts are concentrated on text categorization (Bayer *et al.*, 1996; Schütze and Pedersen, 1994; Schütze *et al.*, 1994), text segmentation (Hearst, 1994; Nomoto and Nitta, 1994), and the application of linguistic techniques to the improvement of the quality of text retrieval systems (Evans and Zhai, 1996; Lewis and Sparck Jones, 1996; Smeaton *et al.*, 1994; Strzalkowski *et al.*, 1994).

Finally, an interesting compromise between in-depth text understanding of documents and standard information retrieval techniques is represented by *information extraction*, a more tractable and robust method which aims at extracting specific types of information from a document by ignoring effectively non-relevant text portions (Hobbs *et al.*, 1994; Jacobs and Rau, 1990; Riloff and Lehnert, 1994). The applied techniques are usually template-based and can also be used to define user profiles in personal *information filtering* systems (see (Belkin and Croft, 1992; Höfferer *et al.*, 1994; Maes, 1994; O'Riordan and Sorensen, 1995; Keane *et al.*, 1996)).

3.2. CONCEPT SPACE REPRESENTATION

The aim of concept analysis is to improve the selectivity of a legal thesaurus in that each thesaurus entry is checked in order to decide whether it can be used as precise descriptor. For this purpose we analyze the concept space to capture all distinct connotations, in particular to detect "hidden word senses" which are often enough not noticed during the process of intellectual indexing.

We first create a vector representation of each concept as function of the surrounding contexts of its usage in the underlying legal corpus. The context window can be freely defined on the basis of the number of preceding and following words, e.g. for most of our evaluations we used the symmetric interval of ± 50 words. The word-based interval definition is not an intrinsic limitation of our model. Other interval delimiters like for example sentences are also implemented. The details are reported in (Schweighofer and Scheithauer, 1996). As next step we apply statistical analysis as well as neural networks to the task of structuring the concept space. A comparison of the two methods is given in Sect. 4.1. The final step of the thesaurus refinement process is the intellectual elimination of descriptor clusters with “fuzzy meanings”, leaving only descriptors with appropriate selectivity.

The generation of the vector representation for the individual concepts is performed according to the process model shown in Fig. 2. The document texts included in the legal corpus are first transformed into sequential word lists by the process of tokenization, i.e. the segmentation into individual words. Subsequently, we eliminate all meaningless stop words from the sequential word list to produce a reduced word list which only contains entries that are significant for the connotation of a concept. Such a stop word elimination is only necessary for subsequent statistical analysis, especially for the creation of cluster descriptions. The quality of the result from neural network training is not affected by keeping stop words in the word list. This represents an essential advantage of the neural approach because it eliminates the time-consuming task of designing an adequate stop word list for the domain.

By use of a *lemmatizing module* we convert the reduced word list to a word index which indicates for each entry a list of all postings, i.e. the document number and position in the document. The lemmatizing module replaces the exact string match for the comparison of two words with a more sophisticated morphological analysis which is of particular importance for highly inflective languages like German. It deals with the following morphological phenomena fine-tuned for the German language: inflections, conjugations, suffixes, and vowel-gradation (Winiwarter, 1995).

The necessary input of knowledge about legal terminology is provided by the descriptors of the thesaurus. The user is not restricted to the use of simple descriptors to represent a legal concept but can also make use of synonyms as well as compound descriptors. In analogy to the generation of the word index the documents are automatically indexed on the basis of the descriptors resulting in a descriptor index with postings for the individual descriptor occurrences.

By merging the word and the descriptor index we create a merged index for each descriptor. For this purpose we check each word posting if it lies within the context window of a descriptor occurrence. With this we obtain for each word a list which indicates the numbers of those descriptor occurrences of which the word contributes to the connotation.

Finally, according to the *vector space model* of information retrieval (Salton and McGill, 1983; Turtle and Croft, 1992) for each descriptor occurrence i a corresponding concept vector C_i is calculated which captures the connotation as function of the presence or absence of certain words in its context, also referred to as its properties:

$$C_i = (WORD_{i1}, WORD_{i2}, WORD_{i3}, \ldots, WORD_{in}) \ . \tag{5}$$

We make use of *binary indexing* for the calculation of the components of the vector, i.e. $WORD_{ik} = 1$ if the word k is part of the context of descriptor occurrence i and $WORD_{ik} = 0$ otherwise.

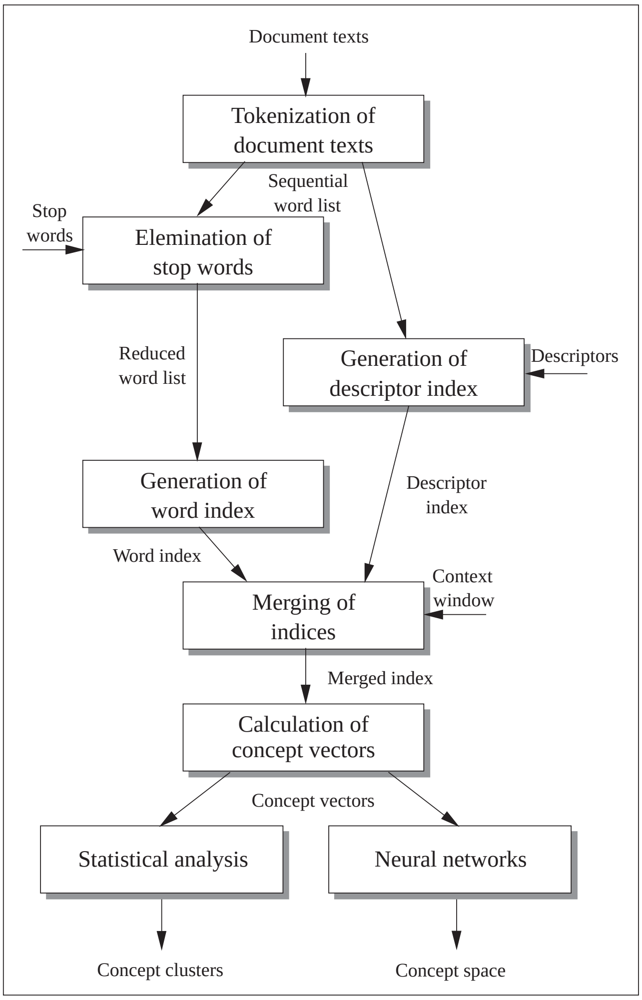


Figure 2. Generation of descriptor vectors.

The resulting concept vectors represent the input for the neural network to compute the concept spaces for the individual concepts. Besides this, we also make use of the statistical analysis as follows. First, for each descriptor we calculate the similarities between its different connotations. In particular, the similarity between two different occurrences of a descriptor is expressed by the number of words that are present in both contexts, i.e. the number of vector components that equal 1 in both descriptor vectors. For this purpose we apply the symmetric *similarity coefficient of Dice* (Salton and McGill, 1983), which indicates the percentage of words that the two contexts have in common. Based on these similarity values, the descriptor occurrences are then clustered by using a *quick partition algorithm* as outlined in Fig. 3. According to a pre-defined threshold for the similarity value, the algorithm produces non-hierarchical disjunctive clusters (see also (Panyr, 1987) or

```
for i := 1 to n do                                          for each descriptor occurrence i
    for j := i+1 to n do                                    compare with all other descriptor occurrences
        if sim[i, j] > t then                               if similarity value exceeds threshold
        begin                                               then
            ready := FALSE;                                 reset flag for successful insertion into cluster
            for k := 1 to m do                              for each cluster
                if (i in c[k]) and (j in c[k]) then         if both descriptor occurrences already in cluster
                    ready := TRUE;                          then do nothing
                else
                    if (i in c[k]) then                     if only i element of cluster
                        begin                               then
                            insert(j, c[k]);                insert j into cluster
                            ready := TRUE;
                        end
                    else
                        if (j in c[k]) then                 if only j element of cluster
                        begin                               then
                            insert(i, c[k]);                insert i into cluster
                            ready := TRUE;
                        end
            if NOT ready then                               if no insertion occurred
                create_clust(c, i, j, m);                   then create new cluster with i and j
        end
merge_clusters(c, m);                                       merge clusters with identical entries
add_singletons(c, m, n);                                    add all missing singleton clusters
```

Figure 3. Quick partition cluster algorithm.

(Jain and Dubes, 1988) for a general exposition of clustering algorithms). Each pair of descriptor occurrences is tested whether its similarity value exceeds the threshold. If this condition is satisfied and one of the two descriptor occurrences is already a member of an existing cluster, then the other one is inserted into that cluster. Only if both of the two descriptor occurrences have not been included in any of the clusters yet, a new cluster is created. In a second run, we verify that there are no clusters with identical entries. If two such clusters are detected, they are merged. Finally, all descriptor occurrences that have not been added to any cluster so far, are appended as singleton clusters.

An important feature of statistical analysis is that the resulting clusters can be easily supplemented by cluster descriptions to obtain meaningful representations of the specific connotation for each cluster. We first retrieve for the cluster members the associated postings from the descriptor index and use then the reduced word list to extract the most frequent words in the concerned context windows. This results in a ranked list of words which provides a very useful and informative output for the interpretation of the result of statistical analysis as well as important information for document analysis (see below).

3.3. DOCUMENT SPACE REPRESENTATION

We make use of the result of statistical concept analysis to produce a representative description of the contents of a legal document (see Fig. 4 for the process model).

In order to rank the descriptor clusters that are included in a certain document according to their relevance, several ranking algorithms were tested (Salton

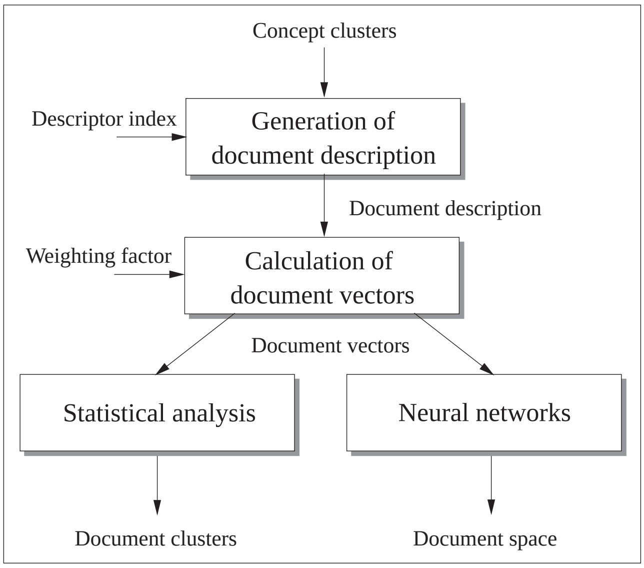


Figure 4. Generation of document description and document space

and Buckley, 1988). We achieved the most satisfying results by using the *inverse document frequency* as weighting factor.

As final result a list of the most relevant descriptor clusters supplemented by the cluster descriptions is displayed for each document to provide an accurate and clear summary of the contents of the document (Schweighofer and Winiwarter, 1993a).

In the following this document representation is also used to calculate the vector representation of a legal document. For each document i a document vector D_i is computed in analogy to the computation of the calculation of concept vectors in Sect. 3.2. The only difference here is that the components of the vector consist of two groups, the descriptors and the words taken from the cluster description:

$$D_i = (DESCR_{i1}, \ldots, DESCR_{im}, WORD_{i1}, \ldots, WORD_{in}) \ . \qquad (6)$$

To emphasize the importance of the descriptors the binary weighting function is modified in that for descriptors a different weighting factor $w > 1$ is applied. Therefore, for descriptors which are contained in the document we set $DESCR_{ik} = w$. With this, the document vectors provide an accurate representation of the contents of a legal document, which combines the use of precise descriptors with terms from the cluster descriptions to indicate specific connotations.

Finally, the document vectors serve as input to a statistical analysis and a neural network module to produce a representation of the structure and the relations in the document space of the legal corpus.

4. Experimental Results

4.1. ANALYSIS OF LEGAL CONCEPT SPACES

As test environment for our approach we used documents from the European Community law database CELEX. The test database for descriptors consists of 41 text segments of documents. The text material, i.e. terms with context windows of ±50 words, was produced as retrieval result from a search in the CELEX database for the term “neutrality”. We selected “neutrality” because this concept is a very good example of a legal term with several meanings. By intellectual separation we achieved clusters of the various context related meanings of the term “neutrality” which represented the comparison module for our automatic analysis, see Fig. 5. Due to space restrictions we can present only the various groups and the CELEX numbers of the documents. Furthermore, each cluster is labeled by a short descriptive term. Note that several segments of one document are designated by using capital letters, e.g. /A, /B, etc.

The efficient clustering algorithm of KONTERM produces sound results. The clusters can be seen as connotations of the concept that are described automatically. A shortcoming of KONTERM is the sensitivity to the correct adjustment of the parameters (i.e. list of stop words, threshold value). However, multiple clustering with different parameters is a useful strategy for the analysis of a term. The result of the clustering algorithm is presented in Fig. 6. For each cluster we give the consecutive number of the text segment as well as its corresponding CELEX number. Furthermore, we provide the cluster description which consists of the ten most frequent words that are contained in the respective contexts.

Comparative experiments were performed with unsupervised neural networks. More precisely, self-organizing maps are trained with the descriptor vectors as input data. The length of these vectors is about 500 components. Geometrically speaking, we perform a projection from a high dimensional input space onto a two-dimensional output space by means of the self-organizing map. The most obvious difference to the statistical approach is that the neural network does not produce clusters but maps. The advantage of such maps is a better description of the relationships between the various connotations of a concept which can be described by using the following geographical terms:

- *Hill*: Strong concentration of document segments with the same meaning,
- *Plateau*: Loose set of document segments with similar meanings,
- *Valley*: Document segments with meaning elements of several groups,
- *Region*: Neighborhood relationship between hills and plateaux.

A note on the graphical representation of the final map which is given in Fig. 7 below is necessary. The graphical representation contains as many entries as there are output units in the artificial neural network. Thus, every entry corresponds to exactly one unit of the self-organizing map. Each entry is further assigned either

- Neutrality of States (STATE): 992E2408, 990H0306, 989H0195, 987H0184, 987H0183, 982H0240
- Fiscal neutrality:
  - Neutrality of the value-added tax system (VAT): 389L0465, 385L0361, 381Y0924(10), 367L0227/A, 367L0227/B, 690C0097, 690C0060, 690C0035
    - deduction of residial VAT (RES-VAT): 689J0159/A, 689J0119
    - import turnover tax (IMP-VAT): 689C0343
    - parent companies and subsidiaries in different member States (SUB-VAT): 390L0435
  - Spirits (SPIRITS): 689C0230
  - CO_2/energy tax (EN-TAX): 392D0180
  - Sugar sector (SUGAR): 390B0354
  - Non-discrimination in matters of taxation (NON-DISC): 689J0159/B, 689J0011/A
- Neutrality of competition:
  - Neutrality of common rules for the allocation of slots at Community airports (SLOTS): 393R0095/A, 393R0095/B, 393R0095/C, 393R0095/D
  - Neutrality of the Community eco-management and audit scheme (ENVIRON): 393R1836, 392R0880
  - Neutrality of the tariff structures in the combined transport of goods (TRANSPORT): 393D0174/A, 393D0174/B, 393D0174/C
  - Neutrality of computer reservation systems for air transport services (AIR-SERV): 391R0083, 388R2672
- Neutrality of the research programs of the Joint Research Centre (RESEARCH): 392D0274
- Neutrality of anti-dumping duties (ANTI-DUMP): 392R0738
- Chemical neutrality:
  - Oil seeds (OIL): 386R2435
- Neutrality of the customs valuation system:
  - Customs value of goods (CUSTOMS): 689J0011/B
- Conjunctural neutrality (CONJUNCT): 385D105.1
- Cost-neutrality (COST): 385D105.3
- Budgetary neutrality (BUDGET): 380Y1231(06)

Figure 5. Intellectual analysis of term “neutrality”.

the CELEX number of a text segment or a dot. The appearance of a label denotes the fact that the corresponding unit exhibits the highest activation level with regard to the input vector corresponding to this CELEX number. Therefore, this unit is the winning unit. On the contrary, a dot appears in the final map if none of the input vectors is assigned to the corresponding unit. In other words, the respective

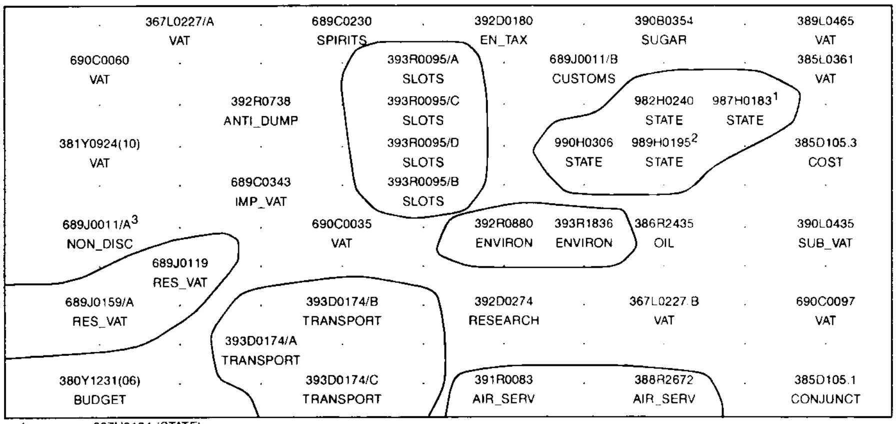
1 41 NEUTRALITY

- /1/11/ 393R1836, 392R0880
  MEMBER, STATES, COMPOSITION, COMPETENT, BODIES, INDEPENDENCE, PROVISIONS, THIS, REGULATION, CONSISTENT
- /2/3/4/ 393D0174/A, 393D0174/B, 393D0174/C
  APPLICATION, TARIFF, STRUCTURE, PRINCIPLE, INCOME, RAILWAY, COMPANIES, CERTAIN, ROUTES, ADJUSTMENTS
- /5/15/18/21/22/27/ 992E2408, 990H0306, 989H0195, 987H0184, 987H0183, 982H0240
  QUESTION, NO, COUNCIL, AUSTRIAN, COMMISSION, BONDE, IRISH, UNILATERAL, DECLARATIONS, WRITTEN
- /6/7/8/9/ 393R0095/A, 393R0095/B, 393R0095/C, 393R0095/B
  MEMBER, STATES, AIRPORT, PRINCIPLE, NO, RESPONSIBLE, TRANSPARENCY, DISCRIMINATION, CERTAIN, REQUIREMENT
- /14/20/ 391R0083, 388R2672
  PARTICULAR, NO, BASIS, CONDITIONS, ORDER, DISCRIMINATION, SUBJECT, REGARDS, SYSTEMS, CO-OPERATION
- /19/30/31/33/ 389L0465, 367L0227/A, 367L0227/B, 690C0060
  TAX, VALUE, ADDED, SYSTEMS, PRODUCTION, DISTRIBUTION, MEMBER, STATES, PROVISIONS, SERVICES
- /28/35/36/37/38/ 381Y0924, 689J0159/A, 689J0159/B, 689J0119, 689J0011
  MEMBER, STATES, TAX, COMPETITION, VAT, REMISSION, EXPORTATION, GOODS, COMMON, RESPECT
- **The other remaining clusters consist of only one single document:**
  392D0274, 392R0738, 392D0180, 390L0435, 390B0354, 386R2435, 385L0361, 385D105.1, 385D105.3, 380Y1231, 690C0097, 690C0035, 689J0011/B, 689C0343, 689C0230

Figure 6. Clusters of term “neutrality”.

unit does not exhibit the highest activation level for any input vector. Due to the limited space in the figures the CELEX number of only one text segment is shown even in the case where more than one text segment is assigned to an output unit. The remaining text segments are given as footnotes. In order to ease comparison we give the short mnemonic description for each CELEX number as they are introduced in Fig. 5. Note that the topological arrangement of the labels may serve as an indication for the similarity of the corresponding text segments. However, the distance of the labels in terms of the two-dimensional surface cannot be used as an exact metric of semantic similarity.

Concerning the final classification result of the neural network we recognize a tight connection between e.g. neutrality of the common rules for the allocation of slots at Community airports and neutrality of states, comparable to the clusters of the statistical analysis. In our more geographically oriented description we might refer to these areas as the “hills” of the map. Furthermore, some highly informative plateaux are formed by e.g. fiscal neutrality or neutrality and environment. To complete this discussion, a region can be seen including the meanings fiscal neutrality, cost neutrality, budgetary neutrality, and conjunctural neutrality.



1. 987H0184 (STATE)
2. 992E2408 (STATE)
3. 689J0159/B (NON_DISC)

Figure 7. Concept space of term "neutrality".

Although the interpretation of the neural network is more sophisticated than the statistical approach, the main advantage remains that the tuning of the model (stop word list, threshold value) is not necessary. Visualization approaches aiming at a more intuitive representation of cluster boundaries are currently a topic of considerable ongoing research, see for instance (Cottrell and Bodt, 1996; Kraaijveld *et al.*, 1992; Merkl, 1997; Merkl and Rauber, 1997b; Merkl and Rauber, 1997a; Ultsch, 1993).

4.2. ANALYSIS OF LEGAL DOCUMENT SPACES

The test environment for document space analysis is a database consisting of 75 full-text documents of court decisions from the European Community law database CELEX. The thesaurus is taken from the lexicon :SUBjects of CELEX which contains some 250 descriptors, more or less corresponding to the major chapters of the treaties and areas of Community activity. Only few descriptors are added to this list. The automatically produced document description is transformed to a weighted document vector (with $w = 9$ for descriptor terms and $w = 1$ for cluster description terms, see Sect. 3.3). This document vector represents the input to the neural network. Some remarks about the quality of the thesaurus are in order. The indexation in the lexicon in CELEX is of average quality because of the low number of descriptors and the stress on the area of Community activity. Although the automatic indexation is paramount to the intellectual one, some inconsistencies remain which can be easily resolved by adding more descriptors. The length of these vectors is about 630 components. As mentioned above, each output unit in the artificial neural network is assigned to the CELEX number and a short mnemonic description. For the sake of clarity we provide the various mnemonic descriptions in Fig. 8 as well.

- Supremacy of Community law (SUPRA)
- Direct applicability of Community law (APPL)
- Direct effect of secondary legislation (EFFECT)
- Questions concerning the European Parliament (seat, conciliation, *locus standi*) (EP)
- Questions concerning the relationship between public international law and Community law (treaty-making power of the European Community, effect of treaties in Community law) (INTLAW)
- Non-contractual liability of the European Community (LIAB)
- Fundamental human rights (RIGHTS)
- Legal base chosen for a legal act (BASE)
- Legal status of regions (REGION)
- Safeguard clauses (SAFE)

Figure 8. Groups of CELEX documents with mnemonic description.

The map as depicted in Fig. 9 shows significant hills concerning the non-contractual liability of the European Community, legal problems of the European Parliament, the relationship between public international law and Community law as well as human rights. A convincing region is formed by the hills concerning direct applicability of Community law and direct effect of secondary legislation. Shortcomings are some "run-aways" which might be due to the poor thesaurus and the merge of descriptors concerning legal questions (e.g. direct applicability of Community law) and areas of Community activity (e.g. agriculture).

5. Conclusion

We described the KONTERM project as an approach towards conceptual legal information retrieval relying on the one hand on linguistic analysis and on the other hand on statistical and connectionist methods to represent similarities in concept and document spaces. More precisely, linguistic analysis is utilized in the sense of exploratory data analysis of text corpora. The goals of exploratory data analysis are two-fold. First, we are interested in establishing precise terms to build up legal thesauri by means of connotational analysis. Second, similarities on the document level are used to classify various documents.

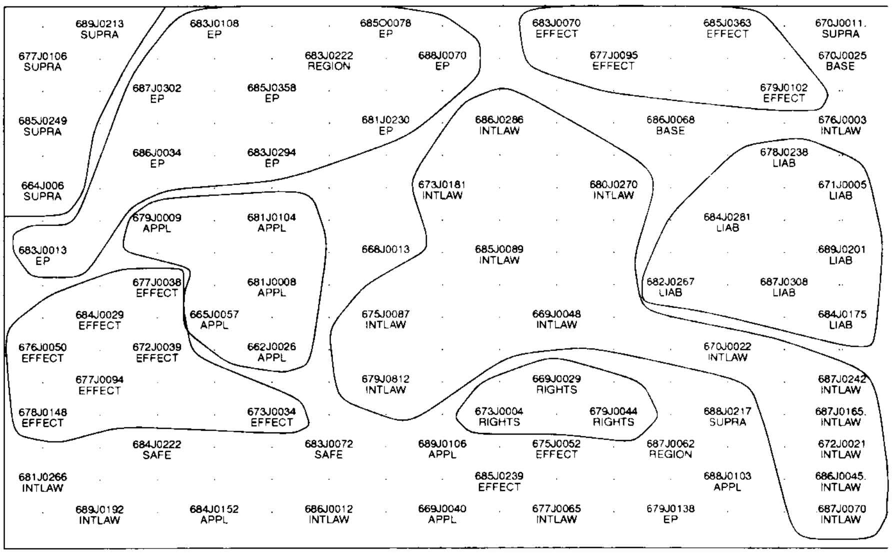


Figure 9. Document space of selected CELEX court decisions

This paper shows that neural networks in combination with statistical methods represent a highly effective tool for the computation of similarities in large text corpora. The statistical method is simple and efficient for a computable approximation of legal texts. Neural network technology is superior to cluster analysis since it proved to be able to produce its results without the need of time-consuming tasks which are related to stop word elimination and threshold selection. Neural networks can automatically produce quite useable maps of descriptor and document spaces.

Future work within the KONTERM project will concentrate on two major issues. First, the input data to the statistical and the connectionist module will be enhanced to comprise syntactic constructs. Additionally, we will make use of a part-of-speech tagger and lexica. The acquired knowledge concerning the legal language during the test phases will be used to fine-tune existing lexica. Another interesting extension might be the automatic analysis of term variations as described in (Jaquemin, 1994). Second, concerning the connectionist approach to concept and document space analysis, we plan to evaluate some of the most recent architectures of unsupervised neural networks. In particular, major emphasis will be directed towards architectures enabling improved and more intuitive visualization of similarities.

Acknowledgments

This research is funded by the *Jubiläumsfonds der Oesterreichischen Nationalbank* (research project no. 4941). We thank the Office for Publications of the European Communities for providing us with the documents that were used to perform the experiments.

Dieter Merkl thanks Ron Sacks-Davis and the people at MDS for providing a perfect working environment during his visit of the Department of Computer Science, Royal Melbourne Institute of Technology; part of this work has been done during his visit.

References

Agirre, E. & Rigau, G. (1996). Word sense disambiguation using conceptual density. In *Int. Conference on Computational Linguistics*.

Allen, J. F. (1987). *Natural Language Understanding*. Benjamin/Cummings, Menlo Park.

Armstrong, S. (ed.) (1994). *Using Large Corpora*. MIT Press, Cambridge, MA.

Ashley, K. D. (1990). *Modeling Legal Argument*. MIT Press, Cambridge, MA.

Bayer, T., Renz, I., Stein, M. & Kressel, U. (1996). Domain and language independent feature extraction for statistical text categorization. In *Workshop on Language Engineering for Document Analysis and Recognition*.

Belew, R.K. (1987). A connectionist approach to conceptual information retrieval. In *Int. Conference on Artificial Intelligence and Law*.

Belkin, N. J. & Croft, W. B. (1992). Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, **35**(12).

Bench-Capon, T. (1993). Neural networks and open texture. In *Int. Conference on Artificial Intelligence and Law*.

Bing, J. (1987). Designing text retrieval systems for conceptual searching. In *Int. Conference on Artificial Intelligence and Law*.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.

Blackmore, J. & Miikkulainen, R. (1995). Visualizing high-dimensional structure with the incremental grid growing network. In *Int. Conference on Machine Learning*.

Bruce, R. & Wiebe, J. (1994). Word-sense disambiguation using decomposable models. In *Annual Meeting of the Association for Computational Linguistics*.

Carpenter, G. A. & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer* **21**(3).

Charniak, E. (1993). *Statistical Language Learning*. MIT Press, Cambridge, Mass.

Church, K. W. & Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, **19**(1).

Church, K. W. & Rau, L. F. (1995). Commercial applications of natural language processing. *Communications of the ACM*, **38**(11).

Cottrell, M. & De Bodt, E. (1996). A Kohonen map representation to avoid misleading interpretations. In *European Symposium on Artificial Neural Networks*.

Cross, G. R. & De Bessonet, C. G. (1985). Representation of legal knowledge for conceptual retrieval. *Information Processing & Management* **21**(1).

Cunningham, H., Wilks, Y. & Gaizauskas, R. J. (1996). New methods, current trends and software infrastructure for NLP. In *Int. Conference on New Methods in Language Processing*.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, **41**.

Dick, J. P. (1991). Representation of legal text for conceptual retrieval. In *Int. Conference on Artificial Intelligence and Law*.

Evans, D. A. & Zhai, C. (1996). Noun-phrase analysis in unrestricted text for information retrieval. In *Annual Meeting of the Association for Computational Linguistics*.

Fritzke, B. (1994). Growing cell structures: A self-organizing network for unsupervised and supervised learning. *Neural Networks* **7**(9).

Fritzke, B. (1995). Growing grid: A self-organizing network with constant neighborhood range and adaptation strength. *Neural Processing Letters* **2**(5).

Fritzke, B. (1996). Growing self-organizing networks: Why? In *European Symposium on Artificial Neural Networks*.

Gelbart, D. & Smith, J. C. (1993). FLEXICON: An evaluation of a statistical ranking model adapted to intelligent legal text management. In *Int. Conference on Artificial Intelligence and Law*.

Greenleaf, G., Mowbray, A. & van Dijk, P. (1995). Representing and using legal knowledge in integrated decision support systems: DataLex WorkStations. *Artificial Intelligence & Law*, **3**(1–2).

Grefenstette, G. (1994). *Explorations in Automatic Thesaurus Discovery*. Kluwer, Boston.

Groendijk, C. & Oskamp, A. (1993). Case recognition and strategy classification. In *Int. Conference on Artificial Intelligence and Law*.

Groendijk, C. (1992). Neural schemata in automated judicial problem solving. In *Legal Knowledge-Based Systems - Information Technology and Law (JURIX'92)*.

Hafner, C. D. (1981). *An Information Retrieval System Based on a Computer Model of Legal Knowledge*. UMI Research Press, Ann Arbor.

Hearst, M. A. (1994). Multi-paragraph segmentation of expository discourse. In *Annual Meeting of the Association for Computational Linguistics*.

Hobbs, J. R., Appelt, D. E., Bear, J. & Israel, D. (1994). FASTUS: A system for extracting information from text. In *ARPA Human Language Technology Workshop*.

Höfferer, M., Knaus, B. & Winiwarter, W. (1994). Adaptive information extraction from online messages. In *RIAO Conference, Intelligent Multimedia Information Systems and Management*.

Honkela, T. Pulkki, V. & Kohonen, T. (1995) Contextual relations of words in Grimm tales analyzed by self-organizing maps. In *Int. Conference on Artificial Neural Networks*.

Jacobs, P. S. & Rau, L. F. (1990) SCISOR: Extracting information from on-line news. *Communications of the ACM* **33**(11).

Jain, A. K. & Dubes, R. D. (1988). *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs.

Jain, A. K. & Mao, J. (1996). Artificial neural networks: A tutorial. *Computer* **29**(3).

Jaquemin, C. (1994). FASTR: A unification-based front-end to automatic indexing. In *RIAO Conference, Intelligent Multimedia Information Systems and Management*.

Jing, Y. & Croft, W. B. (1994). An association thesaurus for information retrieval. In *RIAO Conference, Intelligent Multimedia Information Systems and Management*.

Kangas, J., Kohonen, T. & Laaksonen, J. (1990). Variants of self-organizing maps. *IEEE Transactions on Neural Networks* **1**.

Karov, Y. & Edelman, S. (1996). Learning similarity-based word sense disambiguation from sparse data. Technical Report CS-TR96-05, Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science.

Kaski, S. & Lagus, K. (1996) Comparing self-organizing maps. In *Int. Conference on Artificial Neural Networks*.

Keane, S., Ratnaike, V. & Wilkinson, R. (1996). Hierarchical news filtering. In *Int. Conference on Practical Aspects of Knowledge Management*.

Köhle, M. & Merkl, D. (1996). Visualizing similarities in high dimensional input spaces with a growing and splitting neural network. In *Int. Conference on Artificial Neural Networks*.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics* **43**.

Kohonen, T. (1988). Introduction to neural computing. *Neural Networks* **1**.

Kohonen, T. (1989). *Self-Organization and Associative Memory*. Springer-Verlag, Berlin.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE* **78**.

Kohonen, T. (1993). Generalizations of the self-organizing map. In *Int. Joint Conference on Neural Networks*, Nagoya. Japan.

Kohonen, T. (1995). *Self-Organizing Maps*. Springer-Verlag, Berlin.

Konstantinou, V., Sykes, J. & Yannopoulos, G. N. (1993). Can legal knowledge be derived from legal texts? In *Int. Conference on Artificial Intelligence and Law*.

Kozima, H. & Furugori, T. (1993). Similarity between words computed by spreading activation on an English dictionary. In *Conference of the European Chapter of the ACL*.

Kraaijveld, M. A., Mao, J. & Jain, A. K. (1992). A non-linear projection method based on Kohonen's topology preserving maps. In *Int. Conference on Pattern Recognition*.

Lagus, K., Honkela, T., Kaski, S. & Kohonen, T. (1996). Self-organizing maps of document collections: A new approach to interactive exploration. In *Int. Conference on Knowledge Discovery and Data Mining*.

Lewis, D. D. & Sparck Jones, K. (1996). Natural language processing for information retrieval. *Communications of the ACM* **39**.

Lin, X., Soergel, D. & Marchionini, G. (1991). A self-organizing semantic map for information retrieval. In *Int. ACM SIGIR Conference on Research and Development in Information Retrieval*.

Maarek, Y. S. & Smadja, F. Z. (1989). Full text indexing based on lexical relations – an application: Software libraries. In *ACM SIGIR Int. Conference on Research and Development in Information Retrieval*.

Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM* **37**(7).

Martinetz, T., Berkovich, S. G. & Schulten, K. J. (1993). Neural-gas network for vector quantization and its application to time-series prediction. *IEEE Transactions on Neural Networks* **4**(4).

McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* **5**.

Merkl, D. & Rauber, A. (1997) Alternative ways for cluster visualization in self-organizing maps. In *Workshop on Self-Organizing Maps*.

Merkl, D. & Rauber, A. (1997). On the similarity of eagles, hawks, and cows: Visualization of similarity in self-organizing maps. In *Int. Workshop Fuzzy-Neuro-Systems*.

Merkl, D., Tjoa, A. M. & Vieweg, S. (1992). BRANT: An approach to knowledge based document classification in the information retrieval domain. In *Int. Conference on Database and Expert Systems Applications*.

Merkl, D., Schweighofer, E. & Winiwarter, W. (1994). CONCAT: Connotation analysis of thesauri based on the interpretation of context meaning. In *Int. Conference on Database and Expert Systems Applications*.

Merkl, D. (1995). A connectionist view on document classification. In *Australasian Database Conference*.

Merkl, D. (1995). Content-based document classification with highly compressed input data. In *Int. Conference on Artificial Neural Networks*.

Merkl, D. (1995). The effects of lateral inhibition on learning speed and precision of a self-organizing feature map. In *Australian Conference on Neural Networks*.

Merkl, D. (1997). Exploration of document collections with self-organizing maps: A novel approach to similarity representation. In *European Symposium on Principles of Data Mining and Knowledge Discovery*.

Miikkulainen, R. (1991). Self-organizing process based on lateral inhibition and synaptic resource redistribution. In *Int. Conference on Artificial Neural Networks*.

Minsky, M. L. & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. MIT-Press, Cambridge, MA.

Mital, V., Stylianou, Y. & Johnson, L. (1991) Conceptual information retrieval in litigation support systems. In *Int. Conference on Artificial Intelligence and Law*.

De Mulder, R. V. & van Noortwijk, C. (1994). A system for ranking documents according to their relevance to a (legal) concept. In *RIAO Conference, Intelligent Multimedia Information Systems and Management*.

Ng, H. T. & Lee, H. B. (1996). Integrating multiple knowledge sources to disambiguate word sense; an exemplar based approach. In *Annual Meeting of the Association for Computational Linguistics*.

Nomoto, T. & Nitta, Y. (1994). A grammatico-statistical approach to discourse partitioning. In *Int. Conference on Computational Linguistics*.

O'Riordan, A. and Sorensen, H. (1995). An intelligent agent for high-precision text filtering. In *Int. Conference on Information and Knowledge Management*.

Paice, C. D. (1991). A thesaural model of information retrieval. *Information Processing & Management* **27**(5).

Panyr, J. (1987). Vektorraum-Modell und Clusteranalyse in Information-Retrieval-Systemen (Vector-space model and cluster analysis in information retrieval systems). *Nachrichten für Dokumentation* **38**.

Pao, Y.-H. (1989). *Adaptive Pattern Recognition and Neural Networks*. Addison-Wesley, Reading, MA.

Pereira, F. C. N., Tishby, N. and Lee, L. (1993). Distributional clustering of English words. In *Annual Meeting of the Association for Computational Linguistics*.

Philipps, L. (1989). Are legal decisions based on the application of rules or prototype recognition? Legal science on the way to neural network. In *Pre-Proceedings of the 3rd Int. Conference on Logica, Informatica, Diritto*.

Riloff, E. and Lehnert, W. (1994). Information extraction as a basis for high-precision text classification. *ACM Transactions on Information Systems* **12**(3).

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge.

Ritter, H. & Kohonen, T. (1989). Self-organizing semantic maps. *Biological Cybernetics* **61**.

Rose, D. E. & Belew, R. K. (1989). Legal information retrieval: A hybrid approach. In *Int. Conference on Artificial Intelligence and Law*.

Rose, D. E. (1994). *A Symbolic and Connectionist Approach to Legal Information Retrieval*. Lawrence Erlbaum, Hillsdale.

Rumelhart, D. E. & Zipser, D. (1986). Feature discovery be competitive learning. In Rumelhart, D. E. & McClelland, J. L. (eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition*. MIT-Press, Cambridge, MA.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E. & McClelland, J. L. (eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition*. MIT-Press, Cambridge, MA.

Salton, G. & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management* **24**.

Salton, G. & McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill, New York.

Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Reading.

Schütze, H. & Pedersen, J. O. (1994). A cooccurrence-based thesaurus and two applications to information retrieval. In *RIAO Conference, Intelligent Multimedia Information Systems and Management*.

Schütze, H., Pedersen, J. O. & Hearst, M. A. (1994). Xerox TREC 3 report: Combining exact and fuzzy predictors. In *Text Retrieval Conference*.

Schweighofer, E. & Scheithauer, D. (1996). The automatic generation of hypertext links in legal documents. In *Int. Conference on Database and Expert Systems Applications*.

Schweighofer, E. & Winiwarter, W. (1993a). Legal expert system KONTERM – automatic representation of document structure and contents. In *Int. Conference on Database and Expert Systems Applications*.

Schweighofer, E. & Winiwarter, W. (1993b). Refining the selectivity of thesauri by means of statistical analysis. In *Third Int. Congress on Terminology and Knowledge Engineering*.
Schweighofer, E., Winiwarter, W. & Merkl, D. (1995). Information filtering: The computation of similarities in large corpora of legal texts. In *Int. Conference on Artificial Intelligence and Law*.
Schweighofer, E. (1995). *Wissensrepräsentation in Information Retrieval-Systemen am Beispiel des EU-Rechts (Knowledge Representation in Information Retrieval Systems, An Investigation Based on the Example of EU Law)*. PhD thesis, University of Vienna.
Sergot, M. J., Sadri, F., Kowalski, R. A., Kriwaczek, F., Hammond, P. & Cory, H. T. (1986). The British Nationality Act as a logic program. *Communications of the ACM* **29**(5).
Sharkey, N. (ed.) (1992). *Connectionist Natural Language Processing*. Intellect, Norwood.
Simitis, S. (1970). *Informationskrise des Rechts und Datenverarbeitung (Information Crisis in Law and Data Processing)*. Müller, Karlsruhe.
Sirosh, J. & Miikkulainen, R. (1993). How lateral interaction develops in a self-organizing feature map. In *IEEE Int. Conference on Neural Networks*.
Sirosh, J. & Miikkulainen, R. (1994). Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics* **71**.
Smeaton, A. F., O'Donnell, R. & Kelledy, F. (1994). Indexing structures derived from syntax in TREC-3: System description. In *Text Retrieval Conference*.
Smith, J. C., Gelbart, D., MacCrimmon, K., Atherton, B., McClean, J., Shinehoft, M. & Quintana, L. (1995). Artificial intelligence and legal discourse: The Flexlaw legal text management system. *Artificial Intelligence & Law* **3**(1–2).
Strzalkowski, T., Carballo, J. P. & Marinescu, M. (1994). Natural language information retrieval: TREC-3 report. In *Text Retrieval Conference*.
Susskind, R. E. (1987). *Expert Systems in Law*. Oxford University Press, Oxford.
Tapper, C. (1982). *An Experiment in Use of Citation Vectors in the Area of Legal Data, Complex 9/82*. Universitetsfortlaget, Oslo.
Tokunaga, T., Iwayama, M. & Tanaka, H. (1995). Automatic thesaurus construction based on grammatical relations. In *Int. Joint Conference on Artificial Intelligence*.
Turtle, H. R. & Croft, W. B. (1991). Evaluation of an inference network-based retrieval model. *ACM Transactions on Information Systems* **9**(3).
Turtle, H. R. & Croft, W. B. (1992). A comparison of text retrieval models. *Computer Journal* **35**.
Turtle, H. R. (1995). Text retrieval in the legal world. *Artificial Intelligence & Law* **3**(1–2).
Ultsch, A. (1993). Self-organizing neural networks for visualization and classification. In Opitz, O., Lausen, B. & Klar, R. (eds.), *Information and Classification – Concepts, Methods, and Applications*. Springer-Verlag, Berlin.
van Opdorp, G. J., Walker, R. F., Schrickx, C., Groendijk, C. & van den Berg, P. H. (1991). Networks at work: A connectionist approach to non-deductive legal reasoning. In *Int. Conference on Artificial Intelligence and Law*.
Wan, W. & Fraser, C. (1994). Multiple Kohonen self-organizing maps: Supervised and unsupervised formation with application to remotely sensed image analysis. In *Australian Conference on Neural Networks*.
Werbos, P. J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioural Sciences*. PhD thesis, Harvard University.
Werbos, P. J. (1994) *The Roots of Backpropagation – From Ordered Derivatives to Neural Networks and Political Forecasting*. John Wiley & Sons, New York.
Wermter, S. (1995). *Hybrid Connectionist Natural Language Processing*. Chapman and Hall, London.
Winiwarter, W. (1995). MIDAS — the morphological component of the IDA system for efficient natural language interface design. In *Int. Conference on Database and Expert Systems Applications*.