Human Memory Models and Term Association

Gerda Ruge
Sietec Systemtechnik LIN 12
Otto-Hahn-Ring 6
D-81739 München
Gerda.Ruge@mch.sni.de

Abstract: Results of cognitive psychology research are analysed to explain why it is difficult for retrieval system users to bring to mind alternative search terms. A human memory model is modified in such a way that it produces additional search terms instead of human associations. A small experiment shows that such a spreading activation network can find alternative terms - with a performance similar to the normally used similarity measures.

1. Introduction

A theme that has been worked on in IR for decades is the finding of semantically similar terms for a given search term, e.g. synonyms, broader terms and so on, for supporting a user during search or for automatic query expansion (e.g. /Sparck-Jones 71/). Pure statistic techniques of corpus analysis have not been proven successful (/Ruge 92 or 94 chapter 4/). These statistical approaches usually consist of co-occurrence analysis of terms in documents, abstracts, titles, sentences or text windows. My own work on this topic has shown that semantically similar terms can be found by means of linguistically based corpus analysis (/Ruge 92/). /Grefenstette 94, p. 137 ff./ examined the applicability of such linguistically based alternative terms in the context of Information Retrieval: They can be used as a basis for building up a thesaurus and are even useful to some extent for automatic query expansion.

In this study I would like to discuss a cognitive aspect of the same topic: the human association capability. It is certainly not interesting here for simulating human memory, but for explaining why it is difficult for humans to associate semantically similar terms for improving retrieval results. The origin of these difficulties may give hints on how to conceptualize an automatic term association in such a way that it has a better than human capability in the special aspect of finding semantically similar terms. It may provide a chance to find an approach with a better performance than /Ruge 92/ and /Grefenstette 94/.

SIGIR'95 Seattle WA USA© 1995 ACM 0-89791-714-6/95/07.$3.50

2. A Well Known Memory Model

A memory model that is frequently referred to is /Collins, Loftus 75/, one of the first that was represented as a spreading activation network. I am going to introduce it here briefly in order to explain the spreading activation principal. However, we will later see that it does not adequately explain human word recall capability.

From a static point of view a **spreading activation network** is a graph with weighted arcs. Beneath their names the nodes are associated with numeric values, their **activation** (not shown in fig. 1). Normally first one or more nodes are initialized by a value greater than zero; these nodes represent a description of a problem. The network - its topology and the weights of the arcs - represent the knowledge that is needed for solving the problem. The process of solving a problem consists in propagating the activation starting from the initially activated nodes along the arcs. The so-called **activation function** determines the new activation of a node depending on the actual activations of its neighbours and on the weights of their common arcs. This is iterated. Each iteration of the activation causes a transition of the network to a new state. One tries to design the whole system in such a way that an equilibrium state is reached after a certain number of transitions. The nodes that are then highly activated represent the solution of the problem.

/Collins, Loftus 75/ try to explain various experimental results in the field of human semantic processing. The main properties of their memory model are listed below:

- Nodes are named by words, but represent concepts.
- The arcs represent connections between similar concepts and also between properties. Collins and Loftus assume further node and arc types e.g. word to concept, but these are not shown in figure 1.
- Similar concepts are connected by higher weights than concepts and their properties.
- Humans make decisions on the basis of thresholds in the network, e.g. the decision such whether fire-engines are red.

Collins and Loftus are able to explain several experimental results on the basis of their memory model. Mostly different response times of subjects for several problem types were interpreted. For all these problems the semantic categorization of certain words was necessary or was the main part of the problem. Thus the Collins and Loftus model shows its strengths in explaining the processing of semantic knowledge, especially in recognizing categories - and that was their aim. It is less convincing if it is applied to pure recall, i.e., to the human association capability. In particular, it does not explain why it is difficult for searchers to bring to mind semantically similar terms for improving queries. Conversely, their model predicts that this must be very easy, because of the high weights of arcs of semantic similarity. The memory model discussed in the next section on the other hand deals with human recall capability and explains the mentioned phenomenon.

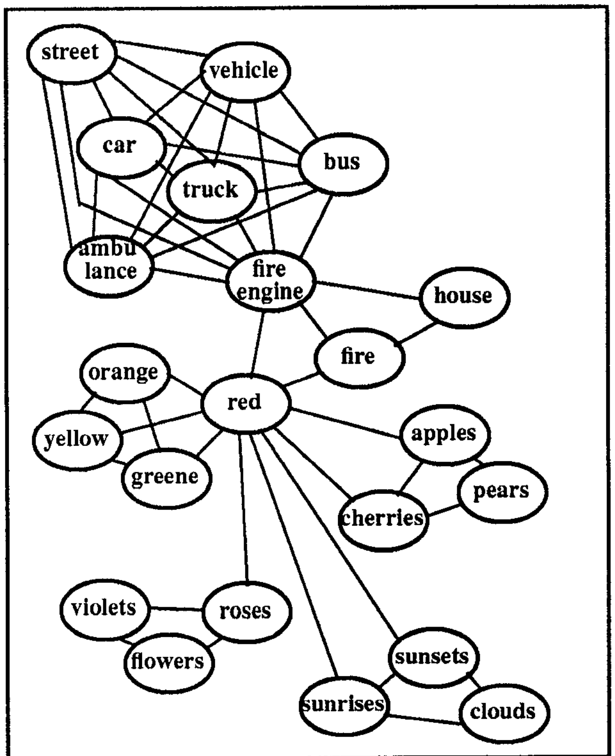

Figure 1: Memory Model of /Collins and Loftus 75, Fig. 1/.

3. A Memory Model Explaining Human Recall Capability

/Wettler 89/ works on automatic term association for finding additional keywords - as I do. But his aims are different to mine in two ways:

- Associations are not generated for a single word, but for a given query, i.e., for a set of words of different meaning.
- The semantic similarity of words is not of primary importance, but rather the simulation of a professional searcher, i.e., a person with special competence in keyword association.

His model is based on the theory of **Classical Associationism**; Wettler formulates its kernel by two laws:

1) The law of associative learning
If two events are in parallel in consciousness, they are linked associatively.

2) The law of associative recall
If an event is in consciousness, all associatively linked events are recalled to consciousness also.

According to the first law, frequent co-occurrence of two words in texts means that they are associated by the reader. Because of the second law, one can assume that human association determines the sequences of words in speech or language production. Therefore strongly associated words are used together during writing - and co-occurrence data from texts reflect human association.

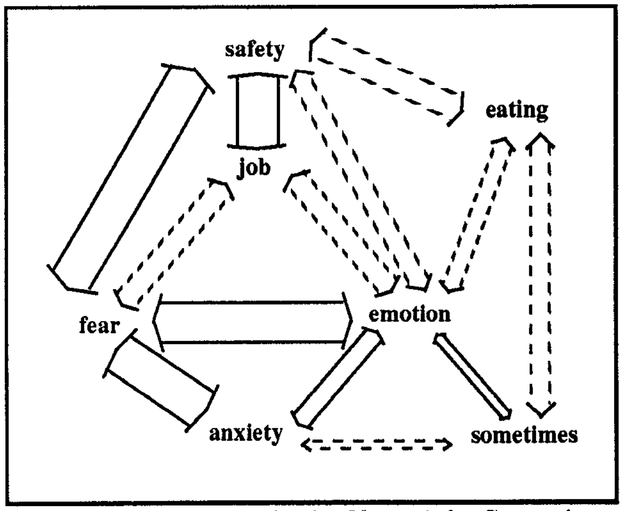

Figure 2: Spreading Activation Network for Generating Alternative Terms (/Wettler 89, Fig. 3/)

Wettler implemented co-occurrence data as a spreading activation network and quantifies the weights of the arcs depending on the probability of two words appearing together or alone in a corpus (/Wettler, Rapp 89/). These weights are symmetric and can be positive or negative. Figure 2 shows such a network; the weights are symbolized by the width of the arcs. Inhibitory arcs are shown in broken lines. For generating additional keywords, the network is used in the following way: First all meaningful words of the query are "set conscious", i.e., they are activated. After some iterations of the activation propagation the high activated nodes are "remembered words"; they can be interpreted as additional search terms. For example if effect, television and aggression are activated the additional terms influence, violence, abuse and behaviour are found. This result seems very natural, thus the simulation of the professional searcher is well done in this case. However the results have not been evaluated.

The approach of Wettler - from a technical point of view - is nothing other than a statistical corpus approach differing only in its computational method. But his approach can be interpreted as a human memory model: It explains the origin of the arcs between the word nodes or conceptual nodes; it is a quantitative model and therefore testable; and it explains the difficulties of human searches in finding alternative search terms: Synonyms do not necessary appear together in the same text, sentence or text window. The arcs between semantically similar terms in the network of Wettler will normally have small weights.

4. Associationism and Term Association

In 1965 Jones already showed the relation between the theory of Classical Associationism and the term statistics in the field of IR (/Jones 65/). According to /Church, Hanks 90/ there are even observable similarities of content: Their results of co-occurrence statistics on the basis of a few word long text windows seemed as if they were produced by free association. Indeed the generated associations contained words that are frequently recalled by free association for some sample words.

The observations of Church and Hanks were tested by /Wettler et al. 93/ in an experiment. They generated term associations on the basis of 12 words long text windows for comparing them with human associations. The data of the comparison were collected some decades before. A single word, the **stimulus**, had been presented to the subjects and the first word that came into their minds had been their response. The statistical associations were very similar to the most frequent human responses - even more similar than the human associations to each other. Wettler's result verifies that co-occurrence is an essential principal of human association. Now I am going to discuss other experiments that show that semantic similarity is not such a principal.

Originally only single association responses were collected from a number of persons in order to establish a norm. This acted as an aid to determine unnormal reactions in single patients (/Miller 93, p. 183/). Nowadays there is also an interest in the variety of associations of single persons as a way to examine the structure of human memory. /Strube 84/ carried out experiments on so-called **free continued association** using the following method: A stimulus is given by name to the subject. He then says whatever words come to his mind without being interrupted. A typical association sequence of Strubes data is shown in figure 3 - with the associations of one subject on one axis and time on the other.

Strube observed that the following is typical: First the subject mentions a few semantically similar words. Then groups of words are mentioned - quickly one after the other with breaks between the groups. The words in the groups are not semantically similar, but "belong to the same situative context" (/Strube 84, p. 75/). He concludes that the associative production of humans is essentially orientated in situative contexts; and the breaks between the groups mean context changes. He points out that situative contexts fit only for explaining the human recall capability, but not for the human judgement capability (p.118); therefore properties would be essential. The main point for this paper is: Only very few associations at the beginning of the chain are words semantically similar to the stimulus. (This is especially a problem for IR).

Beneath the content characteristics of the groups of sequentially rapidly given associations Strube also examined the relations between pairs of following words. He distinguished five different relations (p. 64ff.):

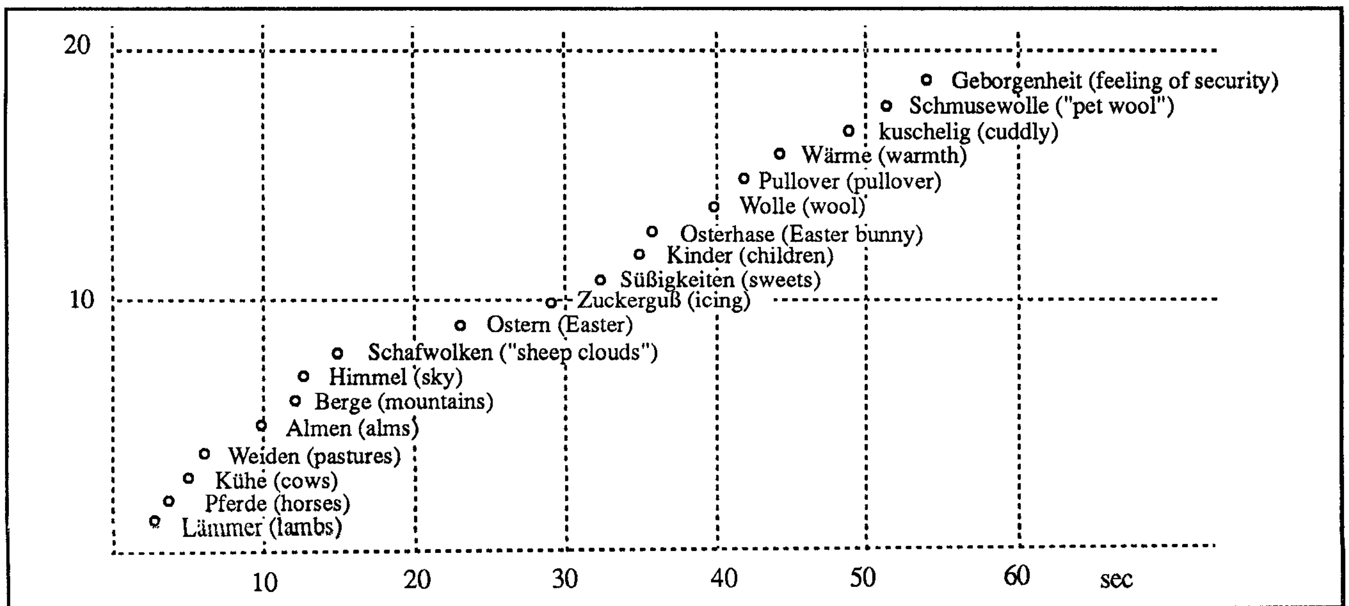

Figure 3: Association Chain with Stimulus <u>Sheep</u> (<u>Schaf</u>) (/Strube 84, p. 68/)

a) paradigmatic relations, normally at the beginning of the association chain, e.g. dirty - unclean

b) fixed concepts, e.g. blue - sky

c) syntagmatic relations, e.g. pork - eat

d) co-occurrence in typical situations, e.g. steak - potato

e) personal / episodical relations, e.g. Woody Allen - white. This production was not understandable. Such associations could be identified as co-occurring in a personal episode of the subjects.

The cases b) - e) are explainable by the two laws of the Classical Associationism, here the words or their concepts obviously co-occurred. I cannot understand, why Strube proposes a model on the basis of features. He points out as a disadvantage of networks, that it is unnecessarily difficult to determine similarities. I think an adequate model must be "difficult" with respect to similarities, that means it should not contain similarity as principal of memory. New results for some special cases show that case a) - synonyms - are also based on the association principal co-occurrence. Let us now have a look at that point.

A long time ago /Deese 64/ found out that stimulus adjectives which are frequent in a general corpus normally provoke their antonyms as response, e.g. large and small and the other way around. This is not the case for rare adjectives, they are associated with common head nouns[1]. His hypothesis was that frequent antonyms frequently co-occur with contextual overlap - as in examples 1 and 2. With this supposition the association of antonyms of frequent adjectives is also explained by the principals of associationism.

example 1: Under normal circumstances, he had a certain bright-eyed all-American-boy charm, with great appeal for **young ladies**, **old ladies**, and dogs.

example 2: Frequently he must work long hours in the **hot sun** or **cold rain.**

example 3: Plato feels that man has two competing aspects, his **rational faculty** and his **irrational.**

The examples 1 - 3 are taken from /Justeson, Katz 91/; they origin from a general corpus of 25,000,000 words. By means of this corpus Justeson and Katz confirmed the Deese hypothesis: Antonymous adjectives which are first place associations responses for each other co-occur in the same sentence significantly more frequently than by chance. In most cases these occurrences also showed syntactic and/or lexical parallels, as in example 1 and 2. This relation also holds for adjectives that are not in Deese's list frequent adjectives, but not as strong[2]. There was also a similar result for rare adjectives whose antonyms are morphological derivates. Here the word pair relation was not tested for single pairs, but - because of their rareness - for the whole class of such pairs.

Justeson and Katz interpret their results as proof that human association is not only established by co-occurrence for syntagmatic relations, but also for paradigmatic relations. Therefore they view associations that are in paradigmatic relation not longer as exceptions to the laws of Classical Associationism.

The discussion so long is now comprised to two hypothesis:

- The essential principal of human association is co-occurrence - even for paradigmatic relations. For words this complies with co-occurrence in text.
- A spreading activation network which is build up on the basis of co-occurrence data is an adequate model for a part of the human association capability.

I am now going to introduce a similar memory model which is linguistically based. Later this model is modified for associating semantically similar terms.

5. Spreading Activation with Heads and Modifiers

The relation between two words that refer to each other in a sentence or phrase is called **head/modifier relation** or simply **link.** The **modifier** is the word that specifies the **head;** in the first sentence of this paragraph two, sentence and phrase are modifiers of the head words. The following example shows the most frequent modifiers and heads of the word storage in patent abstracts (read data storage, heat storage...; storage tank, storage device...):

example 4: storage

modifiers	**heads**
data	tank
heat	device
energy	container
charge	position
water	chamber
information	unit
main ...	battery ...

Pure statistic corpus analysis showed the superiority of short contexts (/Ruge 94, chapter 4/). The shortest meaningful contexts of a word are its heads and modifiers - they consist only of one word. /Ruge 92/ showed that terms with a great overlap in their heads and modifiers are often semantically similar. And /Grefenstette 94/ did similar experiments not only based on the head/modfier relation but also on the verb/subject and verb/object relation. Evidence for the hypothesis that heads and modifiers also fit better than other contexts in a spreading activation memory model can be found in the experimental results discussed so far:

- Human associations of adjectives are often head nouns according to /Deese 64/.

1: The **head** is the word to which the adjective refers.

2: Deese's article does not make clear if he took these words into account or not.

- The contextual overlap that causes the association of antonymous adjectives in human mind are often heads (/Justeson, Katz 91/); certainly not only the same nouns, but semantically similar ones (hot sun - cold rain).
- /Church, Hanks 90/ and /Wettler et al. 93/ simulated human association successfully on the basis of some words' long contexts. One would suppose that these short contexts contain a lot of head/modifier relations.

A further argument for the cognitive relevance of heads and modifiers is the following: For generating the syntactic tree of a sentence humans need special training and explanations of certain grammatical concepts, but every native speaker is able to decide which word refers to which other word in a sentence - and that is the head/modifier relation. Therefore the head/modifier relation is natural in human language.

5.1 Spreading Activation on the Basis of Heads and Modifiers

Figure 4 shows a part of a network based on heads and modifiers - they were extracted from a real world corpus. The function of this network is now explained by an example. First we assume that the weights of all arcs are the same. They must be smaller than 1 because the activation of each node will be spread to its neighbours.

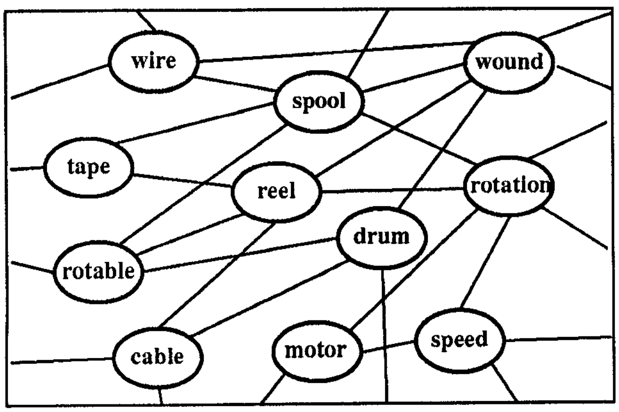

Figure 4: A Part of a Head/Modifier Network

If spool is activated as the initial word, first activation is spread to its heads and modifiers wire, wound, rotation, rotable and tape. With the next propagation iteration, semantically similar words of spool are reached, that is reel and drum, but also words that are not semantically similar like motor. The heads and modifiers of the initial word certainly have an higher activation than the semantically similar words; and further iterations will make little difference. Agreeing with the observations of human association, heads and modifiers of the initial word are the most activated nodes. This was certainly not our aim. We are not looking for a system that can simulate human association capability, but for one that produces semantically similar terms. The network has to be modified in such a way that semantically similar terms are highly activated and heads and modifiers are not.

5.2 Indirect Activation of Semantically Similar Words

The desired effect can easily be realized by defining all arc weights negatively and additionally allowing negative activation. During activation propagation the negative weights do not then effect inhibition but indirect excitation: The heads and modifiers of the initial word are activated negatively; their heads and modifiers however are activated positively. Therefore words sharing heads and modifiers with the initial word are finally activated positively.

Figure 5 shows such a network with spool as initial word after the second iteration of activation propagation. Positively activated nodes are shaded; a darker shade means more activation; negative activated nodes are all white. The estimated activations suppose that all weights are the same with a value between -1 and 0. The total amount of the negative activation of wound and wire - that they got from spool - is greater than the positive one that both exchange with each other. Therefore wire and wound stay with a negative activation, even though they are connected by a negative arc. A similar case is motor and speed. Both are positively activated, certainly less than reel and drum which do not have inhibitory arcs with a positively activated node.

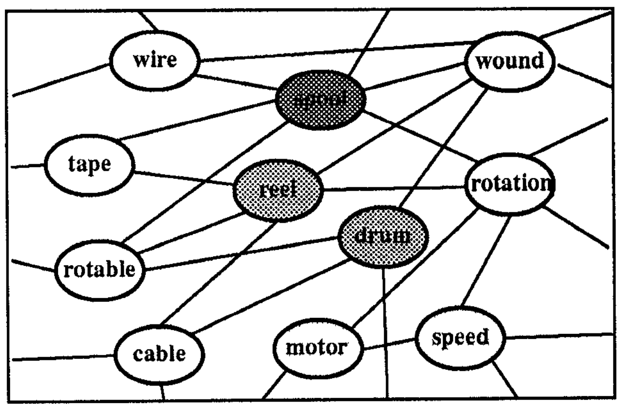

Figure 5:Head/Modifier Network with Indirect Activation

The activation of the heads and modifiers of the initial term is more or less undertunneled by indirect activation: The total amount of activation of the heads and modifiers is greater than those of semantically similar terms, but they are no output terms, because they are negative. Output terms are only the nodes with the greatest positive activation: After two or more iterations these are those words which share many heads and modifiers with the initial term. As shown in /Ruge 92/ these are indeed semantically similar terms.

5.3 Taking into account Synonymous Heads and Modifiers

Until the second iteration, the network does nearly the same as a similarity measure. The essential difference with similarity measures in the vector space model shows its effect after more iterations. Let us look at wire and cable in figure 5. Wire is a modifier of spool, but not of reel and drum. Wire and cable are strongly semantically similar and therefore should be dealt with as overlapping modifiers of spool, reel and drum. The main criticism of the vector space model is that there is no similarity measure that can perform like this (/Raghavan, Wong 86/).

This problem is solved by a spreading activation approach, because negative activation is also propagated. Along the following four arcs long ways the negative activation of wire reaches the node cable:

wire - spool - tape - reel - cable
wire - spool - rotable - reel - cable
wire - spool - rotable - drum - cable
wire - spool - wound - reel - cable
wire - spool - wound - drum - cable
wire - spool - rotation - reel - cable

After four iterations cable receives additional activation from wire. Now it is the case that spool is receiving activation from wire and in parallel reel and drum are getting activation from cable. In the network shown in figure 5 this effect would be very weak. The activation that cable is receiving from wire is much smaller than the one it recieves from other nodes, because the distance between wire and cable is four arcs long and the weights are smaller than 1.

One would assume that wire and cable share many heads and modifiers because they are semantically similar (not shown in figure 5). These heads and modifiers would effect that there is a similar activation for both nodes - along a two arcs long way. Therefore not only the initial word and its semantically similar words have a similar activation - a high positive one - but all similar words in the network receive a similar activation. By means of this mechanism synonymous heads and modifiers have a notable effect on the comparison and as do similar heads and modifiers to some extent.

The main characteristics of the introduced head/modifier network approach can be summarized as follows:

- The network can be built fully automatically by means of linguistic analysis.
- The nodes represent words and the arcs between them the head/modifier relation.
- All weights of the arcs are negative; negative activation is allowed. This results in the new technique of indirect activation.
- The topology of the network is inspired by a human memory model. Words that are normally high activated are undertunneled by the technique of indirect activation. In this way semantically similar words are highly activated.

Such a network has the following advantages and disadvantages with respect to similarity measures:

- Because head arcs and modifiers arcs are not distinguished, the network allowes less fine distinctions than similarity measures.
- The network is more robust.
- Only within the network approach semantic similarity of the heads and modifiers is taken into account.

6. Experiments

The spreading activation approach has advantages and disadvantages with respect to similarity measures, which leads to the question with which approach better results are expected. Both approaches have been implemented in order to compare them. The examination of the network approach could not be carried out as extensively as for the similarity measures in /Ruge 92/. Essentially it was tested wether the network behaved as predicted or not and wether there are obvious differences between the two approaches.

6.1 Test Data

By means of an automatic linguistic analysis all head/modifier relations were extracted from 6,000 abstracts about motors (electric motors and combustion engines) from three annual editions of the US patent office. The linguistic analysis overlooks 15% of the link token and generates about 15% wrong links - usually for types that occur only once. Therefore only those link types were considered that appeared at least twice. The network contained all words of the maximal set of words with ten or more heads or modifiers in this set. That were 1,058 words with in the average 27.5 heads and also 27.5 modifiers.

6.2 Parameters of the Network

The aim of the study was to compare the two approaches spreading activation network and similarity measure in the vector space model. In a first trial the parameters of the network were chosen in such a way that there was no further difference between the network and the similarity measure in /Ruge 92/. The network was defined in such a way that the activations after the second iteration conformed to the similarity values.

In this network the sum of the total amount of activation of all nodes was increasing permanently. This resulted in the following scenario: The most frequent words received the highest activation - independent of the initial word. This network realized a constant function. Therefore it is not possible to take over all parameters of the similarity measure technique and only change the computing technique.

Not all possible network parameters[1] will be discussed here, because they were not examined systematically. Only one example of a spreading activation network is shown here, one that behaves as assumed. Only the main parameters that affect that the network works are discussed, other parameters are only listed. The following parameters describe a network that on the one hand produces semantically similar words and on the other hand takes over as many parameters of the most successful similarity measure from /Ruge 92/ (cosine measure with logarithmic link frequency weight) as possible:

a) The range of the activation is not limited.

b) The weights w_{ij} of the arcs between two nodes a_i and a_j are defined as

$$w_{ij} = \frac{h_{ij}^{ln} + m_{ij}^{ln}}{\sqrt{\sum_{k=1}^{n} h_{ik}^{ln} + m_{ik}^{ln}} \cdot \sqrt{\sum_{k=1}^{n} h_{kj}^{ln} + m_{kj}^{ln}}} \quad (1)$$

where h_{ij}^{ln} and m_{ij}^{ln} are the logarithmic weighted link frequencies of the head/modifier or modifier/head relations between word *i* and word *j*. *n* is the number of words, actually 1,058.

c) The activation of all nodes is propagated simultaneously.

d) The activation function is defined as

$$a_i(t_v) =$$

$$\alpha \cdot a_i(t_{v-1}) + (1-\alpha) \cdot \sum_{j=1}^{n} w_{ij} \cdot a_j(t_{v-1}) \quad (2)$$

where a_i is the activation of node *i*, t_v is the time point and α is a constant with a value between 1 and 0, in the experiment 0.5.

e) The network is started by assigning one word the activation value 1 and 0 to all other words.

The point e) is unusual for spreading activation networks. Their main advantage is the capability of processing information in parallel. A user searching for alternative terms could activate simultaneously several words as initial terms - surely this would result in a better quality of the associated terms. In our experiment only one node may be activated as initial term for comparing the network with the similarity measure approach.

1: /PDP 88/ gives an overview of different network parameters.

By means of the constant α one can control how fast the state of the network changes. The higher the value for α, the slower the network state changes, because only a part of the activation is spread to the neighbours; the rest of the activations stays at the node[2]. If on the other hand the full activation is spread to the neighbours, after two iterations the initial word would not have more activation than the words sharing many heads and modifiers with it. The result state could shift in that case: It is possible that the further activated nodes do not represent the semantically similar words of the initial word, but the semantically similar words of the set of firstly high activated words. Through some such quick changes to the network state the result can drift away from the desired one. To prevent this effect, one must make sure that the initial term has the highest influence on the network state for a long time. This is obtained by a slow change to the network state.

Formula (1) describes the weights of the arcs. The activation of a node is spread to its neighbours according to the logarithmic weight of the link frequencies. For obtaining a symmetric weight also the link frequencies of the neighbours have to be taken into account (second root under division sign). If the weights were not symmetric, the activation would mainly be spread in the directions of large weights. Then there would be the risk that an equilibrium state is not reached.

6.3 Similarity Measure

During spreading activation processing the heads and modifiers of a word are not distinguishable. Therefore a similarity measure sim' was also defined by making no distinction between heads and modifiers - contrary to the experiments of /Ruge 92/, where head similarity and modifier similarity were treated separately. sim' represents the cosine measure with logarithmic weightened link frequencies.

$$sim'(t_i,t_j) = \frac{\left(h_i^{ln} + m_i^{ln}\right) \cdot \left(h_j^{ln} + m_j^{ln}\right)}{\left\|h_i^{ln} + m_i^{ln}\right\| \cdot \left\|h_j^{ln} + m_j^{ln}\right\|} =$$

$$\frac{\sum_{k=1}^{n} \left(h_{ik}^{ln} + m_{ik}^{ln}\right) \cdot \left(h_{jk}^{ln} + m_{jk}^{ln}\right)}{\sqrt{\sum_{k=1}^{n} \left(h_{ik}^{ln} + m_{ik}^{ln}\right)^2} \cdot \sqrt{\sum_{k=1}^{n} \left(h_{jk}^{ln} + m_{jk}^{ln}\right)^2}} \quad (3)$$

2: During so-called simulated annealing (/Cowie, Guthrie 92/) α would be increased slowly to the value 1 to force the network to an equilibrium state.

where
$$h_i + m_i = (h_{i1} + m_{i1}, h_{i2} + m_{i2}, \ldots, h_{in} + m_{in})$$

6.4 Results

Every 10th word of the data base which was sorted according to head and modifier frequency was taken into account in the experiments - 105 words functioned as initial terms. For each of these terms the similarity value with each of the 1,058 terms was computed and sorted to a resulting ranking list, as shown in the examples in figure 6.

The first 10 associations of each initial term were evaluated intellectually by only a single judge to decide if they were semantically similar or not. Terms judged similar are printed bold in figure 6. The **term precision** of the similarity measure was defined as the average value of similar terms in all examined beginnings of the ranking lists. The term precision of the network approach was determined in the same way for the same initial terms. Here the ranking lists were sorted according to the height of the activation of the nodes after 20 iterations.

axis		aperture	
sim'	network	sim'	network
axis	**axis**	**aperture**	**aperture**
direction	**plane**	portion	axis
shaft	**direction**	secured	mirror
plane	slot	**hole**	plane
movement	**center**	**groove**	**hole**
mounted	aperture	**opening**	**channel**
center	portion	supported	**port**
rotation	member	extending	**gap**
member	position	gap	bore
arm	connection	surface	dimension
fastened		**higher**	
sim'	network	sim'	network
fastened	**fastened**	**higher**	**higher**
mounted	resilient	dependent	**smaller**
secured	**mounted**	**low**	conical
attached	transition	**reduced**	**greater**
supporting	armed	**greater**	**increased**
support	**securing**	**constant**	**high**
mounting	shaped	**high**	**constant**
securing	connected	maximum	lower
fixed	hinge	reducing	proportional
supported	guide	proportional	**low**

Figure 6: Examples of Term Associations on the Basis of the Similarity Measure 3 and on the Basis of the Network

The term precision was almost the same in both cases: 0.306 and 0.307. But the content of the resulting ranking lists were different at the beginning (see figure 6, for example aperture or higher). In the examples above, the initial term is always the first term of the ranking list. The comparison of only the term precision without taking into account "term recall" is an assertion, because the values are determined with respect to a ranking list beginning of fixed length (see /Ruge 92/ where you will find also a more detailed description of the evaluation procedure).

7. Valuation of the Spreading Activation Approach

The theme of the spreading activation network could be persued in any amount of detail, and there are many open questions that must be answered by experiments. The parameters of the similarity measures have been examined in much more detail (/Ruge 92/) than those of the network. I am not sure that there is no network at all that performs better than similarity measures. The advantage of the network might be shown if a set of initial nodes were to be activated simultaniously. This is because networks are able to work on complex information structures in parallel - where similarity measures are not. At first sight it seems as if there is no relevant improvement in using spreading activation networks.

Additionally the technical disadvantages of the network have to be considered: This approach is less efficient and can currently not be utilized for large amounts of data. There is also the question of how a large network would behave. /Huberman, Hogg 87/ showed that phase transitions can appear in connectionist models - small changes can cause a totally different behaviour of the network. Normally, connectionists models ensure certain mathematical properties by training the network. These properties permit prediction of the behaviour of the network (see /Peretto 84/ for Hopfield Networks and /Hinton 92/ for layered neural networks). The network that was introduced in this paper - on the other hand - is build up from head/modifier relations of a real world corpus, therefore it is not possible to adhere to special restrictions. One would suppose that such networks are susceptible for the effects described by Huberman and Hogg because for every corpus different network values are expected - the number of nodes, the average number of arcs per node and so on. As a whole the network approach cannot be prefered to the similarity measures.

This result should not lead to the conclusion that the described network is a failure. The opposite is the case: It was shown that it works. The main point that I wanted to show in this study is the following: The approach to produce semantically similar terms is not only based on linguistic theories and co-occurrence theories but also on the theory of Classical Associationism.

References

1: Church, K.; Hanks, P.: "Word Association Norms, Mutual Information, and Lexicography", Comp. Ling. 16(1), 1990, 22-29

2: Collins, A.; Loftus, E.: "A Spreading-Activation Theory of Semantic Processing", Psychological Review 82(6), 1975, 407-428
3: Cowie, J.; Guthrie, G. & L.: "Lexical Disambiguation Using Simulated Annealing", Proc. of COLING 92, Nantes, Vol. 1, 359-365
4: Deese, J.: "The Associative Structure of some Common English Adjectives", J. of Verbal Learning and Verbal Behavior 3(5), 1964, 347-357
5: Grefenstette, G.: "Explorations in Automatic Thesaurus Discovery", Kluver Academic Publishers, Boston, Dordrecht, London, 1994
6: Hinton, G.: "Wie neuronale Netze aus Erfahrung lernen", Spektrum der Wissenschaft, Nov. 1992, 134-143
7: Huberman, B.; Hogg, T.: "Phase Transition in Artificial Intelligence Systemes", AI 33(2), 1987, 155-171
8: Jones, P.: "Historical Foundations of Research on Statistical Association Techniques for Mechanized Documentation", in Stevens et al. 65
9: Jordan, M.: "An Introduction to Linear Algebra in Parallel Distributed Processing", PDP, Vol. 1, 365-422
10: Justeson, J.; Katz, S.: "Co-occurrence of Antonymous Adjectives and Their Contexts", Comp. Ling. 17(1), März 1991, 1-19
11: Miller, A.: "Wörter: Streifzüge durch die Psycholinguistik", Spektrum Akademischer Verlag, Heidelberg, 1993
12: Rummelhart, J.; McClelland, L. (eds.), "Parallel Distributed Processing", MIT Press, Cambridge, 1986, 8. Auflage (1988)
12: Peretto, P.: "Collective Properties of Neural Networks: A Statistical Physics Approach", Biological Cybernetics 50, 1984, 51-62
13: Raghavan, V; Wong, S.: "A Critical Analysis of Vector Space Model for Information Retrieval", JASIS 37(5), 1986, 279-287
14: Ruge, G.: "Ein Spreading-Activation-Netz zum automatischen Generieren von Thesaurusrelationen" in Schaeder, B.; Rieger, B. (eds.): "Lexikon und Lexikographie" (Proc. of GLDV-Jahrestagung 1990), Georg Olms Verlag, Hildesheim, 1990, 141-148
15: Ruge, G.: "Experiments on Linguistically Based Term Associations", Inf. Proc. & Mangement 28(3), 1992, 317-332
16: Ruge G.: "Wortbedeutung und Termassoziation", Georg Olms Verlag, Hildesheim, Zürich, New York, 1995, forthcomming
17: Rummelhart, D.; Hinton, G.; McClelland, J.: "A General Framework for Parallel Distributed Processing" in PDP, Vol. 1, 45-76
18: Sparck-Jones, K.: "Automatic Keyword Classification for Information Retrieval", Butterworth, London, 1971
19: Strube, G.: "Assoziation: Der Prozeß des Erinnerns und die Struktur des Gedächtnisses", Springer Verlag, 1984
20: Wettler, M.: "Eine psychologische Kritik des Intuitionismus in der Kognitionsforschung", Informatik Fachberichte 203 (KIFS 89), 1989, 282- 298
21: Wettler, M.; Rapp, R.: "A Connectionist System to Simulate Lexical Decisions in Information Retrieval" in Pfeifer, R.; Schreter, Z.; Fogelman, F.; Steels, L. (eds.): Connectionism in Perspective", Elsevier, Amsterdam, 1989, 463-469
22: Wettler, M,; Rapp, R.; Ferber, R.: "Freie Assoziation und Kontiguitäten von Wörtern in Texten", Zeitschrift für Psychologie 201, 1993, 103-112
23: Wittum, G.: "Mehrgitterverfahren", Spektrum der Wissenschaft, Apr. 1990, 78-90