Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 56, No. 1 (2003) 111–135

Hypothesis generation guided by co-word clustering

JOHANNES STEGMANN[1], GUENTER GROHMANN[2]

[1]*Free University Berlin, Medical Library, University Hospital Benjamin Franklin, Berlin (Germany)*
[2]*Free University Berlin, Institute of Medical Informatics, Biometry and Epidemiology, University Hospital Benjamin Franklin, Berlin (Germany)*

Co-word analysis was applied to keywords assigned to MEDLINE documents contained in sets of complementary but disjoint literatures. In strategical diagrams of disjoint literatures, based on internal density and external centrality of keyword-containing clusters, intermediate terms (linking the disjoint partners) were found in regions of below-median centrality and density. Terms representing the disjoint literature themes were found in close vicinity in strategical diagrams of intermediate literatures. Based on centrality-density ratios, characteristic values were found which allow a rapid identification of clusters containing possible intermediate and disjoint partner terms. Applied to the already investigated disjoint pairs Raynaud's Disease – Fish Oil, Migraine – Magnesium, the method readily detected known and unknown (but relevant) intermediate and disjoint partner terms. Application of the method to the literature on Prions led to Manganese as possible disjoint partner term. It is concluded that co-word clustering is a powerful method for literature-based hypothesis generation and knowledge discovery.

Introduction

In a series of papers, *Swanson* has developed the concept of CBD (Complementary But Disjoint) literatures (*Swanson*, 1986; 1988; 1989 a; 1989 b; 1990 a; 1990 b; 1991; 1993). A "source" literature dealing with a main concept may have partner ("target") literatures dealing with other main concepts not mentioned in the source literature (and vice versa). The partners, however, may be brought together by identification of sense-bearing "intermediate" terms mentioned by both, source and target literatures. In a systematic description of this concept, *Swanson* and *Smalheiser* (1997) outlined two procedures to identify possible literature partners: procedure I starts with a source literature and tries to find possible target terms via intermediate literatures; procedure II starts with preselected source and target literatures and tries to find interesting relations between them via detection of intermediate terms. Concept, term discovery process and a supportive interactive software package (Arrowsmith) have also been presented to the library community in a more recent review (*Swanson* and *Smalheiser*, 1999).

Received June 19, 2002.

Address for correspondence:
JOHANNES STEGMANN
University Hospital Benjamin Franklin, Medical Library, 12203 Berlin, Germany
E-mail: johannes.stegmann@medizin.fu-berlin.de

0138–9130/2003/US $ 20.00

Procedures I and II have been termed "open" and "closed" discovery processes, respectively (*Weeber* et al., 2001), the former being well suited to generate hitherto unnotified hypotheses for unsolved problems (e.g. cure or better treatment of a disease), the latter offering a means to - perhaps - rapidly find support for a new idea or hypothesis. We would like to propose the term Swanson Linking (SL) as a name for this kind of literature-based discovery where SL may be defined as finding disjoint literature partners by establishing meaningful links between them using information retrieval from bibliographic databases.

The structured pathway from source to target described in procedure I (*Swanson* and *Smalheiser*, 1997) involves screening (by humans) of possible intermediate terms occuring in titles of source papers. To some extent, terms are statistically preselected: only those terms are subjected to human inspection which occur with greater relative frequency in source titles than in titles from the database MEDLINE as a whole. In a second step, the literature corresponding to selected intermediate terms is retrieved and restricted to papers dealing with a beforehand selected major theme, e.g. dietary factors. The resulting literature set is screened and possible target terms are ranked according to the number of their linkages to the intermediate literature (i.e., the higher the number of intermediate literatures containing the target term in question the higher its rank). The final list is then subjected to human inspection (*Swanson* and *Smalheiser*, 1997). By this procedure it was possible to find connections between Raynaud's Disease and Fish Oils and its pharmacologically active constitutent Eicosapentaenoic Acid (*Swanson*, 1986), between Migraine and Magnesium (*Swanson*, 1988), and between some other disjoint literature pairs (*Swanson*, 1990b; *Smalheiser* and *Swanson*, 1996a; 1996b; 1998). Other researchers applied statistical manipulations to all steps of the discovery process with the hope to find candidate terms on top of ranking lists (*Gordon* and *Lindsay*, 1996; *Gordon* and *Dumais*, 1997; *Lindsay* and *Gordon*, 1999). They easily identified high-ranking intermediate terms of interest. e.g. Blood Viscosity in the Raynaud's Disease literature set (*Gordon* and *Lindsay*, 1996; *Gordon* and *Dumais*, 1997) and Epilepsy in the Migraine literature set (*Lindsay* and *Gordon*, 1999). However, applying the same statistics to the intermediate literatures, the already known (by *Swanson*'s work) target terms Fish Oil, Eicosapentaenoic Acid, Magnesium could not be found directly in higher ranks. Magnesium, for example, was found in a high rank only after combining the terms derived from several intermediate literatures, a step which already had been shown as crucial by *Swanson* (1988).

Recently, *Weeber* et al. (2000; 2001) applied concepts, derived from automatic mapping of the text (titles, abstracts) of source literatures by means of the Unified Metathesaurus Language (UMLS), to the SL-based discovery process. Selection of

specific intermediate terms (from the source literature) and of specific target terms (from intermediate literatures) were made after restriction to (intellectually) pre-selected broad semantic concepts. The authors could establish the paths Raynaud's Disease – Fish Oil and Migraine – Magnesium. This application of semantic filters led to some clustering process especially of the intermediate terms. However, the already known intermediate and target terms did not have high ranks based on occurrence frequencies, and term lists had to be screened involving human expertise to select the relevant terms. Again, Magnesium escaped direct detection from analysis of a single intermediate literature and could only be found by combining the results of three different pathways (*Weeber* et al., 2001), i.e., by taking into account the number of links between the target (Magnesium) and the source (Migraine) (*Swanson*, 1988).

Database tomography (DT) is a content analysis tool which identifies literature themes and their relationships in (full text) document sets by virtue of its built-in phrase frequency and phrase proximity algorithms (*Kostoff* et al., 1998). Its usage for finding CBD literatures was outlined by *Kostoff* (1999). The latter paper also contains a synopsis and systematic description of different starting points and pathways to literature-based discovery and innovation within the frame of the CBD concept (*Kostoff*, 1999). DT was reviewed by *He* (1999).

In our approach, we use the keywords (mainly the Medical Subject Headings, MeSH) assigned to documents indexed in MEDLINE and try to map the literatures analysed according to the co-occurrence strengths of the keywords. Co-word analysis was developed for mapping the structure and dynamics of scientific and technological knowledge domains (*Callon* et al., 1986), and may be a powerful tool for knowledge discovery (*He*, 1999). Maps of co-word clusters can be generated which position the clusters according to their internal ("density") and external ("centrality") link strengths (*Callon* et al., 1991). Our hypothesis is that at least some of the "interesting" intermediate and target terms may be found at definite regions of the cluster maps and/or exhibit some other common feature(s) which may be expressed in terms of centrality and density. We test this hypothesis using the two CBD literature pairs Raynaud's Disease / Fish Oils and Migraine / Magnesium already described (*Swanson*, 1986; 1988). The results are then applied to the literature pair Prions / Manganese which – as to our knowledge – has not yet been investigated by the SL method but was recently visualised using co-citation techniques (*Chen* et al., 2001).

Methods

Online retrieval

Online searches were performed in PubMed, the web version of MEDLINE, using the US National Center for Biotechnology Information's (NCBI) retrieval system Entrez. Unless otherwise stated, title searches were performed. Retrieved document sets were downloaded in MEDLINE formate (all fields) from PubMed using its SAVE function.

Term extraction

The terms contained in the RN field (Enzyme Commission Numbers, Chemical Abstracts Service Registry Numbers) and in the MH field (Medical Subject Headings, MeSH) were extracted (without subheadings) from the documents contained in a set and stored in a file. Each line in that file consisted of a unique term and the MEDLINE identifier (field UI) of the document to which the term had been assigned. Terms occuring only once (in one document only) were omitted from the subsequent analysis.

Co-word analysis

Co-word analysis were performed essentially as described (*Callon* et al. 1991; *Coulter* et al., 1998) and comprised several steps. All possible term pairs of a document set were identified and their co-occurrence strengths were determined by calculation of the Equivalence Index (E_{ij}, see below). Only term pairs with $E_{ij} \geq 0.050$ (in some cases: 0.030) were used for subsequent clustering. All term pairs included were then sorted according to decreasing Equivalence Indices. The clustering process started with the term pair of the highest link strength. This pair was then deleted from the total list. In the next step, all pairs containing one of the two terms already clustered were extracted from the total list and sorted according to decreasing Equivalence Indices. Again, the term pair with the highest Equivalence Index (consisting of one term already clustered and one new term) was chosen, the new term was added to the cluster, and the pair was deleted from the total list. These steps – extracting pairs containing one of the terms already clustered, sorting them according to decreasing equivalence indices, choosing the term pair of highest link strength and adding its new term to the cluster – were repeated until the cluster was filled. Then, the next cluster was started by choosing the term pair with the highest equivalence index from the remaining total list. An upper

threshold of ten terms per cluster was set. When the links containing one of the terms already clustered were exhausted prior to filling up the cluster, it contained less than ten terms. Only clusters with three (in some cases: two) or more terms were included into the subsequent analysis.

Measures and definitions

Source, Intermediate, Target literature and term(s): we follow the conventions of *Swanson* (1997). Source literature means the literature set (retrieved by corresponding source terms) dealing with "a question or problem area of scientific interest" (*Swanson*, 1997); here, we deal with literature on diseases as source literature. Target literature means the literature set (retrieved by corresponding target terms) dealing with previously unconsidered possible answers for the problems described in the source literature. Intermediate literature means the literature set (retrieved by corresponding intermediate terms) dealing with concepts present in both, source and target literatures.

Equivalence Index E_{ij}: calculated according to the formula

$$E_{ij} = (C_{ij})^2 / (C_i \times C_j) ,$$

where C_{ij} is the number of co-occurrences of terms i and j (i.e., the number of documents in which terms i and j co-occur), and C_i and C_j are the numbers of occurrences of term i and term j, respectively (*Callon* et al., 1991). The coefficient E_{ij} is also called Strength, *S* (*Coulter* et al., 1998).

Density: cluster density was measured by calculating the mean of its internal Equivalence Indices (strenghts of the internal links, i.e., links between cluster members) (*Turner* et al., 1988; *Coulter* et al., 1998).

Centrality: cluster centrality was measured by calculating the mean of its external Equivalence Indices (strenghts of the external links, i.e., links from cluster members to terms contained in other clusters) (*Cambrosio* et al., 1993).

Centrality-Density Ratio (CDR): quotient of cluster centrality and density (*Courtial* et al., 1993).

Source-Intermediate Ratio (SIR): ratio of CDR_{source} and $CDR_{sintermediate}$, where CDR_{source} denotes the CDR value of the source cluster (containing the main source term) and $CDR_{sintermediate}$ denotes the CDR value of an intermediate cluster (containing an intermediate term) derived from the source literature set. The higher CDR value was divided by the lower one, i.e., if $CDR_{source} \geq CDR_{sintermediate}$ then $SIR = CDR_{source}/CDR_{sintermediate}$, otherwise $SIR = CDR_{sintermediate}/CDR_{source}$. This resulted in SIR values equal to or above 1.

Target-Intermediate Ratio (TIR) : ratio of CDR_{target} and $CDR_{tintermediate}$, where CDR_{target} denotes the CDR value of the target cluster (containing the main target term), and $CDR_{tintermediate}$ denotes the CDR value of an intermediate cluster (containing an intermediate term) derived from the target literature set. As for SIR, the higher CDR value was divided by the lower one.

Source-Target Ratio (STR) : ratio of $CDR_{isource}$ and $CDR_{itarget}$, where $CDR_{isource}$ denotes the CDR value of the source cluster (containing the main source term), and $CDR_{itarget}$ denotes the CDR value of the target cluster (containing the main target term), both derived from the same intermediate literature set. Again, the higher CDR value was divided by the lower one.

Graphical display

Cluster sets are graphically displayed using standard software in so-called strategical diagrams (*Callon* et al., 1991) according to their cluster centrality (abscissa) and density (ordinate). The axes' origin is determined by the medians of centrality and density.

Programming

Self-made Perl (version 5.004) scripts were used for term extraction, calculation of equivalence indices, clustering process and calculation of density and centrality values. Scripts were run – by remote connection – on a Unix machine with IRIX (version 6.5.12f), supplied by the computer center of the Free University Berlin.

Results

From the source literature on Raynaud's Disease to the target terms Eicosapentaenoic Acid, Arginine, and Nitric Oxide

Figure 1 shows the distribution of clusters derived from the Raynaud's Disease literature set which consists of 801 documents (publication years 1966 to 1985) all carrying the term Raynaud in the title and comprises 464 unique MeSH or RN terms (search profiles, number of extracted terms and other details of cluster analyses are given in the Appendix). Clusters containing potential intermediate terms (Blood Viscosity, Platelet Aggregation, Blood Platelets etc.) which already have been shown to lead from the source (Raynaud's Disease) to the target (Fish Oil) literature (*Swanson*,

1986) seem to be concentrated in the lower left quadrant of the diagram (i.e., exhibit below-median centrality and density values). A similiar distribution is exhibited in a centrality-density diagram of clusters derived from the Fish Oil literature set (Figure 2). However, possibly interesting intermediate terms are found in other parts of the diagrams, too, and not each cluster in the lower left quadrant contains an interesting intermediate term. Nevertheless, from Figures 1 and 2 one may draw the conclusion that it makes sense to start cluster checking for possible interesting intermediate terms in the lower left quadrant of a strategical diagram of source or target literatures.

To find out if there is a more direct path leading to interesting intermediate term we tried to numerically relate the source cluster (cluster containing the main source term Raynaud's Disease) as well as the target cluster (cluster containing the main target term Eicosapentaenoic Acid, the active compound of Fish Oil) to intermediate clusters of their literature sets by calculating Source-Intermediate Ratios (SIR) and Target-Intermediate Ratios (TIR) (see Methods). Table 1 lists some of these SIR and TIR values. It is certainly difficult to say that there is a "preferred number"; however, in an open discovery process (*Weeber* et al., 2001) it might make sense to start cluster checking with a SIR value around 2 (analyzing the source literature) or a TIR value around 1.5 (analyzing the target literature). In addition, the special kind of the clustering process (see Methods) tends to accumulate intermediate terms of possible interest, e.g. cluster 15 of the Raynaud's Disease literature set holds four relevant terms, and cluster 23 of the Fish Oil literature set holds three relevant terms (see Table 1).

Next, we analysed three intermediate literatures. Figures 3 a, 4, and 5 show the distribution of the clusters derived from the literature (title searches) of Blood Viscosity, Platelet Aggregation and Thrombosis, respectively. In these examples we find the clusters containing the source term Raynaud's Disease and the target term Eicosapentaenoic Acid in (more or less) close neighbourhood. We also determined Source-Target Ratios (STR) (see Methods) in the intermediate literature sets, the target being the cluster containing the term Eicosapentaenoic Acid. The STR values are roughly equal having values slightly above 1 (Table 2), i.e., the source and the target cluster within an intermediate literature set show roughly the same centrality-density ratio (CDR). Analyzing the clusters resulting from a search for the MeSH term BLOOD VISCOSITY, we find the clusters containing the source and the target term in close neighbourhood (Figure 3 b), too, their STR value being almost identical to its counterpart derived from the title search (Table 2).

Table 1. Source-Intermediate Ratios (SIR) of the Raynaud's Disease literature set and Target-Intermediate Ratios (TIR) of the Fish Oil literature set

Intermediate term	SIR	TIR
Blood Viscosity	2.03 (15)	1.50 (25)
Platelet Aggregation	2.03 (15)	1.52 (23)
Blood Platelets	2.03 (15)	1.52 (23)
Platelet Function Tests	2.03 (15)	1.45 (20)
Muscle, Smooth, Vascular	2.67 (34)	1.52 (23)
Erythrocytes	1.81 (32)	1.50 (25)
Endothelium	2.67 (34)	1.12 (15)
Prostaglandins	1.46 (19)	1.72 (29)
Blood Pressure	2.27 (39)	1.30 (9)
Autoimmune Diseases	1.90 (31)	1.12 (1)
Necrosis	3.62 (47)	1.40 (10)
Thrombosis	3.10 (41)	n.d. (31)

Determination of SIR and TIR: see Methods.
Numbers in parentheses: number of cluster containing the intermediate term.
n.d.: not determined (only two terms in cluster 31)

Table 2. Centrality-Density Ratios (CDR) of source and target clusters and Source-Target Ratios (STR) in literature sets being intermediates between the source literature Raynaud's Disease and the potential target terms Eicosapentaenoic Acid, Arginine and Nitric Oxide

Intermediate literature	CDR_{source}	CDR_{target}			STR		
	RD	EPA	ARG	NO	EPA	ARG	NO
Blood Viscosity	0.44 (3)	0.43 (12)	–	–	1.02	–	–
Blood Viscosity (MH)	0.42 (159)	0.45 (19)	–	–	1.07	–	–
Platelet Aggregation	0.74 (90)	0.62 (103)	0.86 (122)	0.55 (63)	1.19	1.16	1.35
Thrombosis	0.50 (16)	0.43 (35)	0.40 (56)	–	1.16	1.25	–

Determination of CDR and STR : see Methods.
Numbers in parentheses: number of cluster containing the source term or one of the target terms.
MH: Medical Subject Heading
RD: Raynaud's Disease; EPA: Eicosapentaenoic Acid; ARG: Arginine; NO: Nitric Oxide

In the diagram of the intermediate literature on Platelet Aggregation the positions of cluster 63 (containing the possible target term Nitric Oxide) and of cluster 122 (containing the possible target term Arginine) are labeled, too, both being in the vicinity of the source cluster (see Figure 4). In the Thrombosis intermediate literature, Arginine is also a member of a cluster being not far away from the source cluster (see Figure 5). The STR values of Arginine and Nitric Oxide are in some cases higher than the STR value of Eicosapentaenoic Acid but still well below 1.40. Both terms, Arginine and

Nitric Oxide, have not been mentioned in the SL literature published so far on the Raynaud's Disease / Fish Oil literature pair (*Swanson*, 1986; *Gordon* and *Lindsay*, 1996; *Gordon* and *Dumais*, 1997; *Weeber* et al. 2000, 2001) although they are potent target terms. L-Arginine is the endothelial substrate for Nitric Oxide formation (*Moncada* et al., 1989), and elevation of Nitric Oxide concentration as therapeutic option for patients with Raynaud's phenomen and scleroderma was only recently suggested (*Freedman* et al., 1999). However, according to our knowledge, the first paper mentioning both, Arginine and Raynaud's Disease, appeared only in 1988 (*Kahan*, 1988), and the first paper mentioning Nitric Oxide together with Raynaud's Disease was published only in 1991 (*Agostini* et al., 1991). The possible beneficial effect of Fish Oil or its compound Eicosapentaenoic Acid to Raynaud's Disease is to activate the enzyme nitric-oxide synthase (*Omura* et al., 2001).

From the results described above, one may draw the conclusion that in an open discovery process – having found (by whichever method) relevant intermediate terms – it is a promising path (i) to draw centrality-density diagrams of the term clusters, identify the source term clusters and look primarily in their close neighbourhood for possibly relevant target terms, and/or (ii) to determine the centrality-density ratios of the term clusters, identify the source term clusters and screen carefully especially those clusters having CDR values similiar to the source term clusters (i.e. STR values of around 1).

From the source literature on Migraine to the target terms Magnesium (deficiency) and Valproic Acid

Next, we analysed the Migraine literature which has been shown to be complementary but unconnected to Magnesium (*Swanson*, 1988) and asked whether the co-word clustering method is able to bring Migraine and Magnesium in close connection and to reveal paths between Migraine and Magnesium and possible other target terms. We conducted a PubMed title search for the word Migraine (restricted to publication years 1966 to 1987, see Appendix) and subjected the 2581 documents retrieved to the clustering process described in Methods. Figure 6 shows the centrality-density diagram of this Migraine literature set. The clusters containing possible intermediate terms already mentioned by *Swanson* (1988) – Spreading (Cortical) Depression, Epilepsy, Platelet Aggregation – are concentrated in the lower left quadrant of the diagram (Figure 6), thus confirming the results obtained from the cluster analysis of the Raynaud's Disease literature (see Figure 1). It is interesting that these three terms have similiar source-intermediate ratios (SIR, see Methods) of 3.10 to 3.39 (Table 3).

Table 3. Source-Intermediate Ratios (SIR) of the Migraine literature set

Intermediate term	SIR
Spreading Cortical Depression	3.10 (147)
Epilepsy	3.39 (132)
Platelet Aggregation	3.30 (114)
Hypoxia, Brain	3.10 (147)
Anorexia Nervosa	2.20 (99)
Schizophrenia	2.20 (99)

Determination of SIR: see Methods.
Numbers in parentheses: number of cluster containing the intermediate term.

For detection of possible target terms we analysed the literature of the intermediate terms just mentioned and included further terms: Hypoxia, Brain (member of the same cluster as Spreading Cortical Depression), Anorexia Nervosa, and Schizophrenia (see Figure 6 and Table 3). The two latter MeSH terms were selected due to their SIR values of around 2, and thus could be well used as guides to target terms provided the conclusion drawn from the analysis of the Raynaud's Disease literature (see above and Table 1) can be applied to the Migraine – Magnesium path. The results are shown in Figures 7 to 12 and Table 4. The Migraine source term clusters and the Magnesium target clusters are – with the exception of the Schizophrenia diagram (Figure 12) – in close vicinity (Figures 7 to 11). The corresponding STR values are between 1.01 and 1.42 (again with the exception of the Schizophrenia set) (Table 4) and thus mostly in the range of the STR values of the intermediate literatures connecting Raynaud's Disease with Eicosapentaenoic Acid, Arginine and Nitric Oxide (see Table 2). The only intermediate literature which guides directly to Magnesium Deficiency as a hypothesized causal factor of migraine (*Swanson*, 1988) is Epilepsy, where the MeSH terms Magnesium and Magnesium Deficiency are harbored by the same cluster (Figure 8). Note also that – whereas the intermediate concept Spreading Cortical Depression did not lead to Magnesium using the method described by *Weeber* et al. (2001) – we find Magnesium in very close vicinity to Migraine in the Spreading Cortical Depression diagram (Figure 7). In the Hypoxia diagram (Figure 10) we detected a cluster very close to the source Migraine cluster containing the possible target term Valproic Acid with a STR value of 1.09 (Table 4). We found this term also in the Epilepsy and Schizophrenia diagrams (Figures 8, 12), in the former being as close to Migraine as Magnesium (Figure 8) with a somewhat higher STR value (Table 4), and in the latter as far away from Migraine as Magnesium (Figure 12) with a STR value of 2.03 which is almost identical to the STR value of the Magnesium cluster in the

Schizophrenia set (Table 4). Valproic Acid, not mentioned in the SL literature (*Swanson*, 1988; *Lindsay* and *Gordon*, 1999; *Weeber* et al., 2001), was first discussed in 1988 as "a new drug in migraine prophylaxis" (*Sørensen*, 1988) (starting in 1991, this paper was cited 70 times until 2002 – data not shown) . The prophylactic effect was only recently confirmed (*Kinze* et al., 2001), and Valproic Acid is amongst the drugs recommended by the German Migraine and Headache Society for migraine prophylaxis in childhood (*Evers* et al., 2002).

Table 4. Centrality-Density Ratios (CDR) of source and target clusters and Source-Target Ratios (STR) in seven literature sets being intermediates between the source literature Migraine and the potential target terms Magnesium and Valproic Acid

Intermediate Literature	CDR_{source}	CDR_{target}		STR	
	MIGR	Mg	VA	Mg	VA
Spreading Cortical Depression	0.49 (10)	0.44 (9)	–	1.11	–
Epilepsy	0.49 (133)	0.42 (126)	0.64 (95)	1.17	1.31
Platelet Aggregation	0.76 (182)	0.75 (176)	–	1.01	–
Hypoxia, Brain	0.77 (108)	0.99 (95)	0.71 (141)	1.29	1.09
Anorexia Nervosa	0.60 (72)	0.85 (122)	–	1.42	–
Schizophrenia	0.81 (41)	0.41 (89)	0.40 (38)	1.98	2.03

Determination of CDR and STR: see Methods.
Numbers in parentheses: number of cluster containing the source term Migraine or one of the target terms. MIGR: Migraine; Mg: Magnesium; VA: Valproic Acid

Definition of rules

The results obtained from the analysis of the Migraine source and intermediate literatures confirm the observations made by analysing the Raynaud's Disease literature sets. Interesting intermediate terms can be found in a region of low centrality and low density of a strategical diagram of the clusters derived from the source literature. Some of the clusters with possible intermediate terms seem to show a characteristic Source Intermediate Ratio (SIR, see Methods) which is around 2 in the case of the Raynaud's Disease set (see Table 1) and around 3 for the Migraine set (see Table 3). In the latter, however, we also find promising intermediate terms in clusters whith an SIR of about 2. Starting the analysis with a (known) target literature set (as the Fish Oil set being target literature for Raynaud's Disease, see above) leads also to clusters with a characteristic Target Intermediate Ratio (here around 1.5, see Table 1) containing the same interesting intermediate terms (we have not yet performed a similiar analysis of the Magnesium target literature).

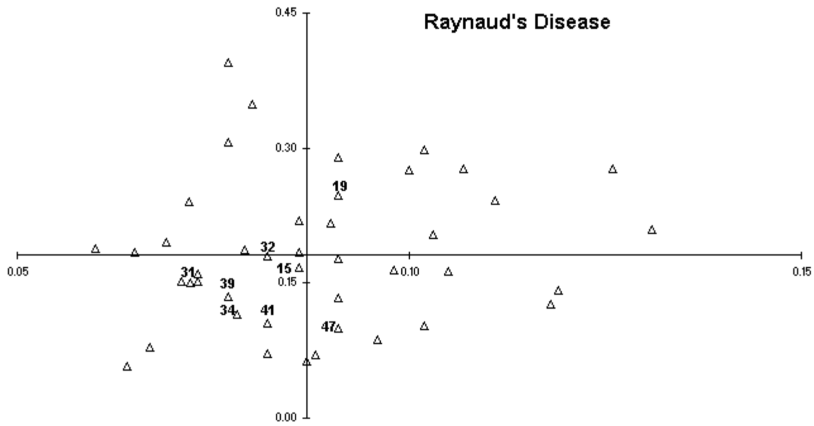

Figure 1. Strategical diagram of the Raynaud's Disease literature set (title search, 1966-1985).
Numbered clusters contain the intermediate terms listed in Table 1.
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.087; median density = 0.182

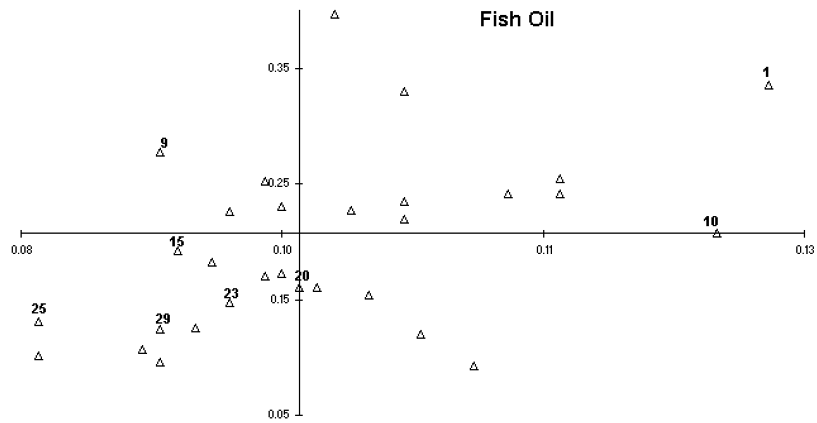

Figure 2. Strategical diagram of the Fish Oil literature set (title search, 1966-1985).
Numbered clusters contain the intermediate terms listed in Table 1.
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.096; median density = 0.208

The cluster analysis of the intermediate literatures has revealed that "unconnected" literature partners can be visualized as "close neighbours" (Figures 3 – 5, 7 – 11).

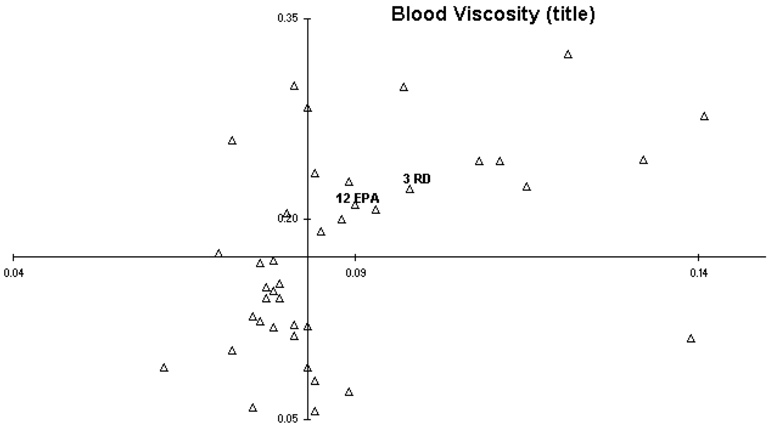

Figure 3a. Strategical diagram of the Blood Viscosity literature set (title search, 1966-1985).
3 RD: cluster 3 containing the source term Raynaud's Disease;
12 EPA: cluster 12 containing the target term Eicosapentaenoic Acid (see also Table 2).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.083; median density = 0.172

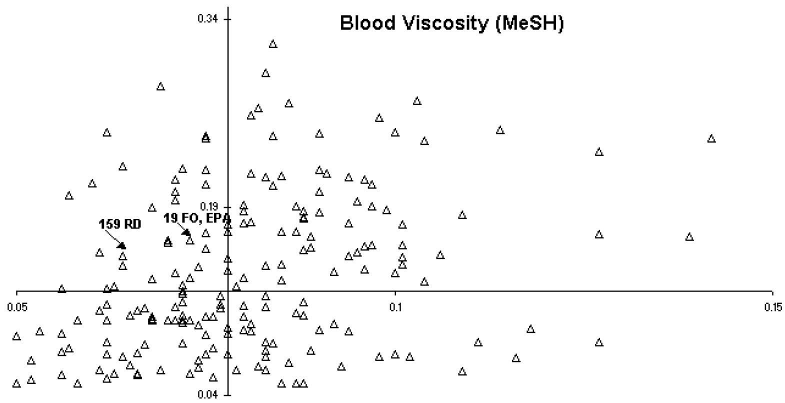

Figure 3b. Strategical diagram of the Blood Viscosity literature set (MeSH search, 1966-1985).
159 RD: cluster 159 containing the source term Raynaud's Disease;
19 FO, EPA: cluster 19 containing the target terms Fish Oils and Eicosapentaenoic Acid (see also Table 2).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.078; median density = 0.123

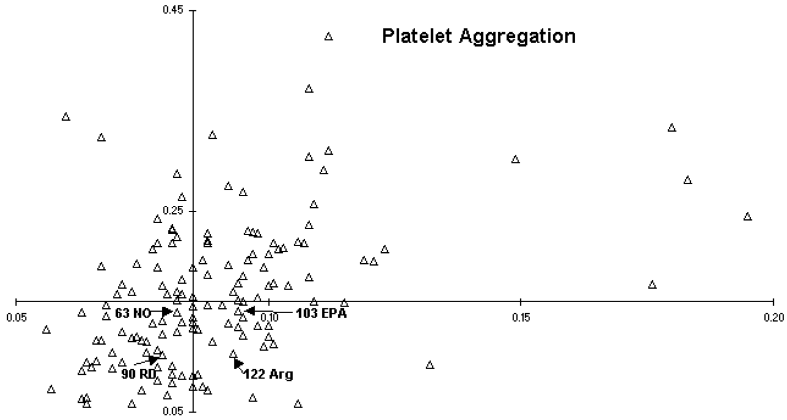

Figure 4. Strategical diagram of the Platelet Aggregation literature set (title search, 1966-1985).
90 RD: cluster 90 containing the source term Raynaud's Disease;
103 EPA: cluster 103 containing the target term Eicosapentaenoic Acid;
63 NO: cluster 63 containing the target term Nitric Oxide;
122 Arg: cluster 122 containing the target term Arginine (see also Table 2).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.085; median density = 0.160

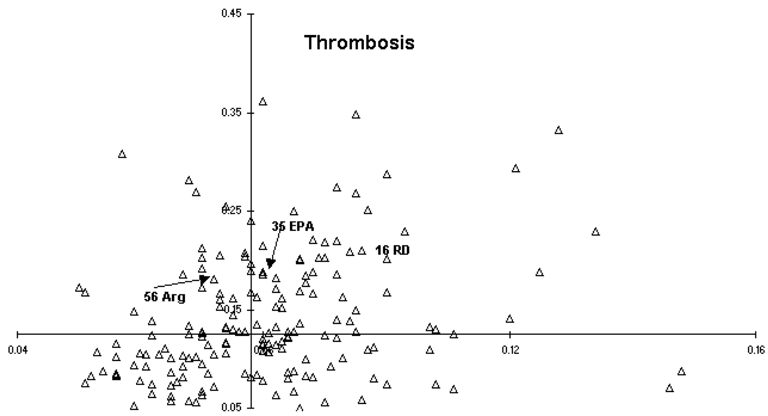

Figure 5. Strategical diagram of the Thrombosis literature set (title search, 1966-1985).
16 RD: cluster 16 containing the source term Raynaud's Disease;
35 EPA: cluster 35 containing the target term Eicosapentaenoic Acid;
56 Arg: cluster 56 containing the target term Arginine (see also Table 2).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.078; median density = 0.125

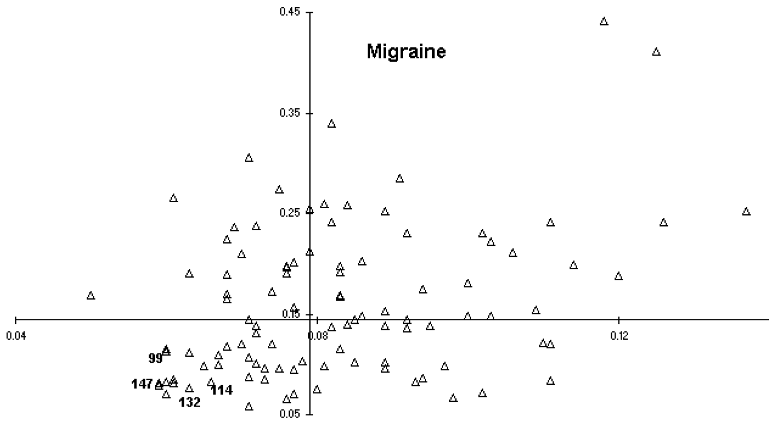

Figure 6. Strategical diagram of the Migraine literature set (title search, 1966-1987).
Numbered clusters contain the intermediate terms listed in Table 3.
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.079; median density = 0.144

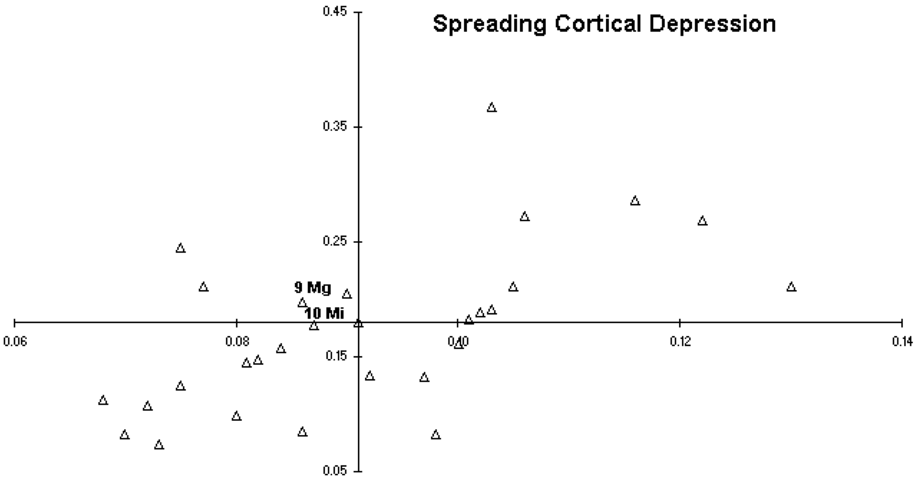

Figure 7. Strategical diagram of the Spreading Depression literature set (title search, 1966-1987).
10 Mi: cluster 10 containing the source term Migraine;
9 Mg: cluster 9 containing the target term Magnesium (see also Table 4).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.091; median density = 0.180

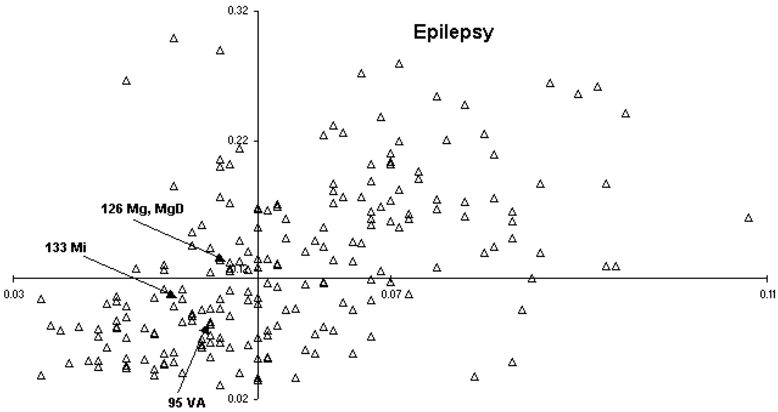

Figure 8. Strategical diagram of the Epilepsy literature set (title search, 1966-1987).
133 Mi: cluster 133 containing the source term Migraine;
126 Mg, MgD: cluster 126 containing the target terms Magnesium and Magnesium Deficiency;
95 VA: cluster 95 containing the target term Valproic Acid (see also Table 4).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.056; median density = 0.114

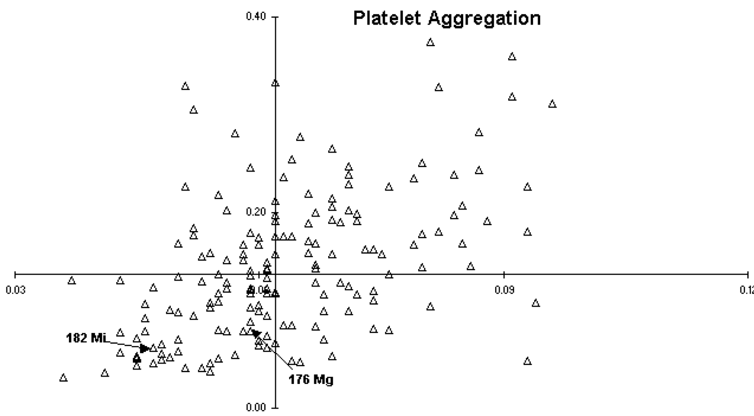

Figure 9. Strategical diagram of the Platelet Aggregation literature set (title search, 1966-1987).
182 Mi: cluster 182 containing the source term Migraine;
76 Mg: cluster 176 containing the target term Magnesium (see also Table 4).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.062; median density = 0.137

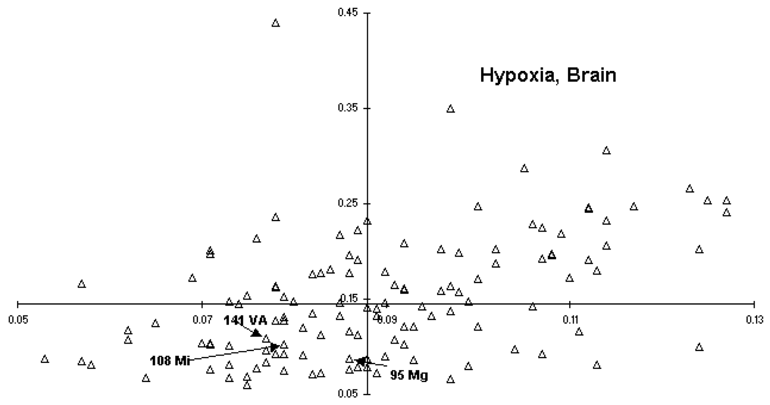

Figure 10. Strategical diagram of the Hypoxia, Brain literature set (MeSH search, 1966-1987).
108 Mi: cluster 108 containing the source term Migraine;
95 Mg: cluster 95 containing the target term Magnesium;
141 VA: cluster 141 containing the target term Valproic Acid (see also Table 4).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.088; median density = 0.145

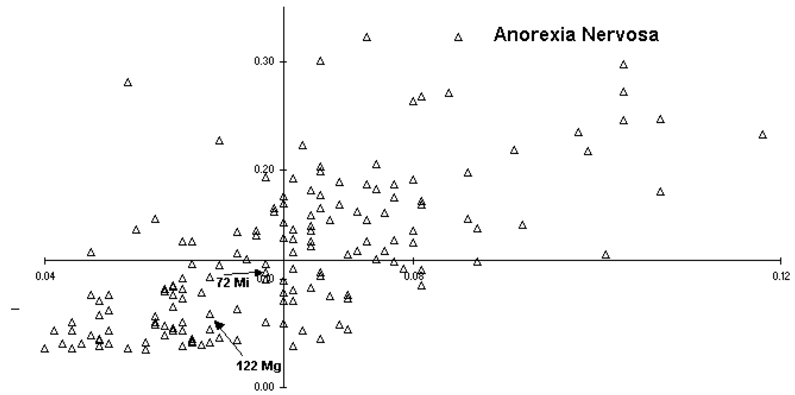

Figure 11. Strategical diagram of the Anorexia Nervosa literature set (MeSH search, 1966-1987).
72 Mi: cluster 72 containing the source term Migraine;
122 Mg: cluster 122 containing the target term Magnesium (see also Table 4).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.066; median density = 0.117

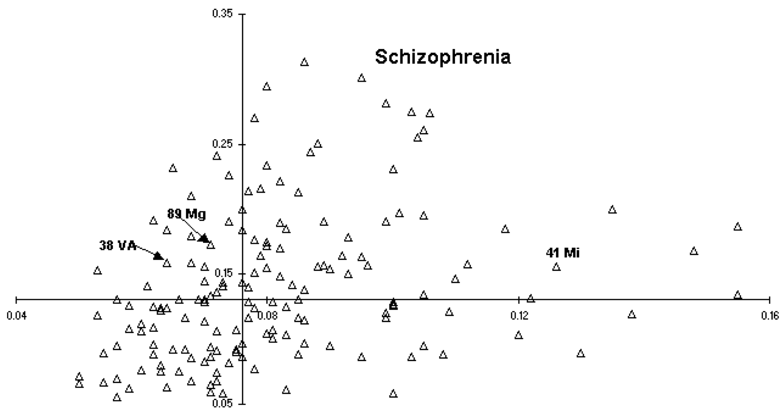

Figure 12. Strategical diagram of the Schizophrenia literature set (title search, 1966-1987).
41 Mi: cluster 41 containing the source term Migraine;
89 Mg: cluster 89 containing the target term Magnesium;
38 VA: cluster 38 containing the target term Valproic Acid (see also Table 4).
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.076; median density = 0.131

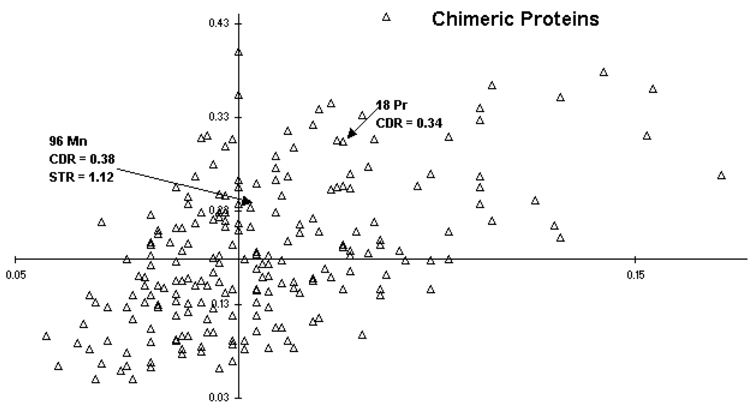

Figure 13. Strategical diagram of the Chimeric Proteins literature set (all fields search, 1966-1995).
18 Pr: cluster 18 containing the source term Prions;
96 Mn: cluster 96 containing the target term Manganese.
Centrality-Density Ratios (CDR) and Source-Target Ratio (STR) are also indicated.
Details of online retrieval and clustering process: see Methods and Appendix.
Median centrality = 0.086; median density = 0.178

Moreover, in several cases Source-Target Ratios (STR, see Methods) show a characteristic value of about 1 (Tables 2, 4). Thus, we may formulate two rules for detecting paths from source to target literatures:

Rule 1: Perform co-word cluster analysis of the keywords (MeSH-, RN terms) assigned to the documents retrieved from a PubMed search. Identify those clusters having below-median centrality and density values. In addition, look for clusters having a source-intermediate ratio (SIR) of about 2 (or 3). Screen these clusters for interesting intermediate terms.

Rule 2: Retrieve the intermediate literature conducting a PubMed title or MeSH search and subject them also to co-word cluster analysis. Identify the clusters containing the main source term and screen the clusters surrounding the source cluster for possible target terms. In addition, screen those clusters having a source-target ratio (STR) of about 1.

We applied these rules to the literature on Prions and, as will be shown in the next paragraph, were able to find a path to the possible target term Manganese.

From the source literature on Prions to the target term Manganese

In a very recent paper, *Chen* et al. (2001) showed that four low cited papers published by *Purdey* (1994; 1996 a; 1996 b; 1998), dealing with organophosphate pesticides being possibly involved in the induction of Bovine Spongiform Encephalopathy (BSE), can be visualised by appropriate expansion of a mainstream-knowledge citation network to "latent" domain knowledge (*Chen* et al., 2001). The four *Purdey* papers mentioned above are the basis for the later explicitly formulated hypothesis that the transformation of the normal cellular prion protein prp_c to its abnormal isoform prp_{sc} is related to manganese concentration (*Purdey*, 2000; 2001). In a thorough literature search for papers dealing with both, Manganese and Prions or Prion Diseases, we retrieved only two papers published prior to 1996; one of them dealt with the cation requirements of RNA viruses (*Scolnick* et al., 1970), the other presented a hypothesis that related multiple sclerosis with geochemical conditions (*Layton* and *Sutherland*, 1975). No direct connections between Manganese and Prions or Prion Diseases (at that time classified as "slow viruses" and "slow virus diseases") were made; however, the idea that heavy metal concentrations are possibly critical for occurrence of this type of mental disorder was already expressed (in the latter article).

Therefore, the Prions/Manganese pair seems to be a good candidate for SL analysis, and we tried to find a direct path from Prions to Managanese using the co-word

clustering method described above. We searched for prions and prion diseases (PubMed title search, restricted to publication years 1966 to 1995, see Appendix), downloaded the 2848 documents retrieved and subjected them to co-word cluster analysis as described in Methods.

From a cluster with a SIR value of 2.03 – located in the below-median centrality and density region of the cluster diagram – we selected the term Chimeric Proteins as intermediate term, conducted a PubMed search (see Appendix) and downloaded the retrieved 2611 documents. In the strategical diagram of this intermediate literature we find the cluster containing the target term Manganese in the wider neigbourhood of the cluster containing the source term Prions. The corresponding STR value is 1.12 (Figure 13), thus confirming the results of the preceding analyses (see Table 2 and 4).

Discussion

Our results verify the hypothesis (see Introduction) that co-word analysis applied to keywords according to *Callon* et al. (1991)) reveals paths between CBD literatures and thus may contribute to discovery of implicit knowledge and serve as an important tool in SL experiments. In strategical diagrams of source or target literatures, clusters containing intermediate terms are found in regions of low density and low centrality (see Figures 1, 2, 6) and show characteristic SIR (Tables 1, 3) or TIR values (Table 1). In strategical diagrams of intermediate literatures, clusters containing source and target terms are found in close vicinity (Figures 3–5, 7–11, 13) and exhibit characteristic STR values (Tables 2, 4). The paths opened by this form of cluster analysis are not exclusive: other diagram regions may also contain interesting terms (see, e.g., Figure 2), and STR values may differ considerably from the "main" value of around 1 (see, e.g., Table 4). However, our results point to a "first choice" of finding intermediate and target (or source) terms when co-word clustering is applicated and analysis of diagrams and clusters is needed. Moreover, we were not only able to achieve the results reported by *Swanson* (1986, 1988), but also found for both source literatures, Raynaud's Disease and Migraine, additional target terms (see Table 2, 4), not mentioned earlier, which were proved to be of therapeutical significance (see Results).

Using co-word clustering within the frame of the SL method we also found evidence of a link between Manganese and Prions (see Results), an almost "classical" CBD literature pair in so far as (i) no direct link between the two partner exists within the publication period analysed, (ii) the "knowledge" of a possible direct connection between the two partners is "implicit", i.e., can be easily derived from the published literature reading the right paper (*Layton* and *Sutherland*, 1975) with a ready mind, and

(iii) later, a well-grounded theory involving the two partners on the basis of field observations has been formulated (*Purdey*, 2000).

Intermediate literature may connect more than one pair of CBD literature partners. For instance, Platelet Aggregation connects Raynaud's Disease with Fish Oil (Figure 4, Table 2) and Migraine with Magnesium (Figure 9, Table 4). Although the two Platelet Aggregation literature sets are not identical (due to different publication windows, see Appendix), we find also in the earlier (1966-1985) Platelet Aggregation literature a cluster containing the source term Migraine and another cluster containing the target term Magnesium, both clusters having similiar CDR values (data not shown). Thus, it could be worthwile to retrieve conjectured intermediate literature without prior selection of source or target terms, subject its keywords to cluster analysis, group the clusters according to their CDR values and inspect clusters with similiar CDR values for possibly complementary source and target terms. Doing so, the preselected intermediate term would be treated as a independent variable (*Kostoff*, 1999).

Of course, human expertise is then necessary for cluster inspection and classification of the contained terms (*Kostoff*, 1999).

Our results show that the "classical" form of co-word analysis employed by us seems to be a powerful method for detection of CBD literature partners and their linking literatures, and major parts of the processes of preselection of intermediate and target terms may be performed automatically. Human intervention is, of course, needed for final selection of appropriate terms, but a priori restriction to broad concepts for selection of intermediates (*Weeber* et al., 2001) or targets (*Swanson* and *Smalheiser*, 1997) is not necessary. In addition, screening of clusters of limited size, containing linked terms (see Methods), seems to be easier than reading of (long) lists of alphabetically sorted or ranked (according to whichever statistics) terms. One major drawback of co-word analysis performed on keywords is its thesaurus dependence, i.e., only documents indexed by virtue of a controlled vocabulary can be analysed. With respect to the examples presented in this communication, the use of MeSH and RN terms (see Methods) is obviously sufficient to achieve correct and encouraging results. However, full text (titles, abstracts) analysis is certainly needed in cases of missing keywords, whether the documents do not yet contain assigned keywords (as in the pre-Medline subset of PubMed), or the whole database does not use a thoroughly controlled vocabulary (e.g. SCISEARCH), or documents retrieved from different databases (with different vocabularies) are analysed together. Here, mapping of texts to concepts – as described by *Weeber* et al. (2000; 2001) – and analysis of phrase frequencies and proximities (*Kostoff* et al., 1998; *Kostoff*, 1999), combined with human expertise are the methods of choice.

Self-made Perl scripts are our present tools for data processing (see Methods). They do their job right, but are far from being perfect. We are currently working on a more "professional" system on the basis of Java and SQL running on a normal PC and controllable within a web browser. However, especially to people with no background in informatics (e.g. librarians), but keen on doing own research, we recommend to acquire some basic Perl programming skills (see also *Katz* and *Hicks*, 1997) and to use, at least initially, their own scripts for text analyses in the context of own research projects on "knowledge discovery". Thus, the challenge brought to the library community by *Swanson*'s work (commented already by *Davies*, 1989) can and should now be accepted by librarians.

Conclusion

We have successfully adopted the known method of co-word analysis as a tool for SL-based hypothesis generation and are now ready to perform further experiments with respect to the structure of CBD literatures as well as the investigation of other literature candidates being possibly CBD partners.

*

A very small and preliminary part of this study has been presented during the annual meeting of the German Association of Medical Libraries, Hamburg, September 17–19, 2001.

We thank Santosh *Nigam*, Eicosanoid Research Division of our institution, for explanation of the metabolic concepts involving arginine as substrate for nitric oxide formation.

We are grateful to the staff of the Free University's computer center for helpful comments on using their services.

References

AGOSTONI, A., B. MARASINI, M. L. BIONDI, C. BASSANI, A. CAZZANIGA, B. BOTTASSO, M. CUGNO (1991), L-arginine therapy in Raynaud's phenomenon? *International Journal of Clinical & Laboratory Research*, 21 : 202–203.

CALLON, M., J. LAW, A. RIP (1986), *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*, London: The Macmillan Press Ltd.

CALLON, M., J. P. COURTIAL, F. LAVILLE (1991), Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry, *Scientometrics*, 22 : 155–205.

CAMBROSIO, A., C. LIMOGES, J. P. COURTIAL, F. LAVILLE (1993), Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis, *Scientometrics*, 27 : 119–143.

CHEN, C., J. KULJIS, R. J. PAUL (2001), Visualizing latent domain knowledge, *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, 31 : 518–529.

COULTER, N., I. MONARCH, S. KONDA (1998), Software engineering as seen through its research literature: a study in co-word analysis, *Journal of the American Society for Information Science*, 49 : 1206–1223.

COURTIAL, J. P., M. CALLON, A. SIGOGNEAU (1993), The use of patent titles for identifying the topics of invention and forecasting trends, *Scientometrics*, 26 : 231–242.

DAVIES, R. (1989), The creation of new knowledge by information-retrieval and classification, *Journal of Documentation*, 45 : 273–301.

EVERS, S., R. PORTHMANN, M. ÜBERALL, E. NAUMANN, W. D. GERBER (2002), Therapie idiopathischer Kopfschmerzen im Kindesalter. Empfehlungen der Deutschen Migräne- und Kopfschmerzgesellschaft (DMKG). [Treatment of idiopathic headache in childhood – recommendations of the German Migraine and Headache Society (DMKG)], *Schmerz*, 16 : 48–56.

FREEDMAN, R. R., R. GIRGIS, M. D. MAYES (1999), Acute effect of nictric oxide on Raynaud's phenomenon in scleroderma, *Lancet*, 354 : 739.

GORDON, M. D., S. DUMAIS (1998), Using latent semantic indexing for literature based discovery, *Journal of the American Society for Information Science*, 49 : 674–685.

GORDON, M. D., R. K. LINDSAY (1996), Toward discovery support systems: a replication, re-examination and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil, *Journal of the American Society for Information Science*, 47 : 116–128.

HE, Q. (1999), Knowledge discovery through co-word analysis, *Library Trends*, 48 : 133–159.

KAHAN, A., H. AWADA, Y. SULTAN, C. J. MENKES, B. AMOR (1988), Tissue plasminogen activator (t-pa) activity and t-pa inhibition (pai) in systemic sclerosis. *Arthritis and Rheumatism*, 31 (Suppl. 4) : S112.

KATZ, J. S., D. HICKS (1997), Desktop Scientometrics, *Scientometrics*, 38 : 141–153.

KINZE, S., M. CLAUSS, U. REUTER, T. WOLFT, J. P. DREIER, K. M. EINHÄUPL, G. ARNOLD (2001), Valproic acid is effective in migraine prophylaxis at low serum levels: a prospective open-label study, *Headache*, 41 : 774–778.

KOSTOFF, R. N. (1999), Science and technology innovation, *Technovation*, 19 : 593–604.

KOSTOFF, R. N., H. J. EBERHART, D. R. TOOTHMAN (1998), Database tomography for technical intelligence: a roadmap of the near-earth space science and technology literature, *Information Processing & Management*, 34 : 69–85.

LAYTON, W., J. M. SUTHERLAND (1975), Geochemistry and multiple sclerosis: a hypothesis, *Medical Journal of Australia*, 1 : 73–77.

LINDSAY, R. K., M. D. GORDON (1999), Literature-based discovery by lexical statistics, *Journal of the American Society for Information Science*, 50 : 574–587.

MONCADA, S., R. M. PALMER, E. A. HIGGS (1989), The biological significance of nitric oxide formation from L-arginine. *Biochemical Society Transactions*, 17 : 642–644.

OMURA, M., S. KOBAYASHI, Y. MIZUKAMI, K. MOGAMI, N. TODOROKI-IKEDA, T. MIYAKE, M. MATSUZAKI (2001), Eicosapentaenoic acid (EPA) induces Ca^{2+}-independent activation and translocation of endothelial nitric oxide synthase and endothelium-dependent vasorelaxation, *FEBS Letters*, 487 : 361–366.

PURDEY, M. (1994), Are organophosphate pesticides involved in the causation of bovine spongiform encephalopathy (BSE)? Hypothesis based upon a literature review and limited trials on BSE cattle, *Journal of Nutritional Medicine*, 4 : 43–82.

PURDEY, M. (1996 a), The UK epidemic of BSE: slow virus or chronic pesticide-initiated modification of the prion protein? Part 1: mechanisms for a chemically induced pathogenesis/transmissibility, *Medical Hypotheses,* 46 : 429–443.

PURDEY, M. (1996 b), The UK epidemic of BSE: slow virus or chronic pesticide-initiated modification of the prion protein? Part 2: an epidemiological perspective pathogenesis/transmissibility, *Medical Hypotheses,* 46 : 445–454.

PURDEY, M. (1998), High-dose exposure to systemic phosmet insecticide modifies the phosphatidylinositol anchor on the prion protein: the origins of new variant transmissible spongiform encephalopathies? *Medical Hypotheses*, 50 : 91–111.

PURDEY, M. (2000), Ecosystems supporting clusters of sporadic TSEs demonstrate excesses of the radical-generating divalent cation manganese and deficiencies of antioxidant co factors Cu, Se, Fe, Zn. Does a foreign cation substitution at prion protein's Cu domain initiate TSE? *Medical Hypotheses*, 54 : 278–306.

PURDEY, M. (2001), Does an ultra violet photooxidation of the manganese-loaded/copper-depleted prion protein in the retina initiate the pathogenesis of TSE? *Medical Hypotheses*, 57 : 29–45.

SCOLNICK, E., E. RANDS, S. A. AARONSON, G. J. TODARO (1970), RNA-dependent DNA polymerase activity in five RNA viruses: divalent cation requirements, *Proceedings of the National Academy of Sciences of the United States of America*, 67 : 1789–1796.

SMALHEISER, N. R., D. R. SWANSON (1996a), Indomethacin and Alzheimer's disease, *Neurology*, 46 : 583.

SMALHEISER, N. R., D. R. SWANSON (1996b), Linking estrogen to Alzheimer's disease: an informatics approach, *Neurology*, 47 : 809–810.

SMALHEISER, N. R., D. R. SWANSON (1998), Calcium-independent phospholipase A_2 and schizophrenia, *Archives of General Psychiatry*, 55 : 752–753.

SØRENSEN, K. V. (1988), Valproate: a new drug in migraine prophylaxis, *Acta Neurologica Scandinavica*, 78 : 346–348.

SWANSON, D. R. (1986), Fish oil, Raynaud's syndrome, and undiscovered public knowledge, *Perspectives in Biology and Medicine*, 30 : 7–18.

SWANSON, D. R. (1988), Migraine and magnesium: eleven neglected connections, *Perspectives in Biology and Medicine*, 31 : 526–557.

SWANSON, D. R. (1989a), Online search for logically-related noninteractive medical literatures: a systematic trial-and-error strategy, *Journal of the American Society for Information Science*, 40 :356–358.

SWANSON, D. R. (1989b), A second example of mutually isolated medical literatures related by implicit, unnoticed connections. *Journal of the American Society for Information Science*, 40 : 432–435.

SWANSON, D. R. (1990a), Medical literature as a potential source of new knowledge, *Bulletin of the Medical Library Association*, 78 : 29–37.

SWANSON, D. R. (1990b), Somatomedin C and arginine: implicit connections between mutually isolated literatures, *Perspectives in Biology and Medicine*, 33 : 157–186.

SWANSON, D. R. (1991), Complementary structures in disjoint literatures. In: A. BOOKSTEIN, Y. CHIARAMELLA, G. SALTON, V. V. RAGHAVAN (Eds), *SIGIR'91: Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval (Chicago, Oct. 13-16)*. New York: Association for Computing Machinery, pp. 280–289.

SWANSON, D. R. (1993), Intervening in the life cycles of scientific knowledge, *Library Trends*, 41 : 606–631.

SWANSON, D. R., N. R. SMALHEISER (1997), An interactive system for finding complementary literatures: a stimulus to scientific discovery, *Artificial Intelligence*, 91 : 183–203.

SWANSON, D. R., N. R. SMALHEISER (1999), Implicit text linkages between Medline records: using Arrowsmith as an aid to scientific discovery, *Library Trends*, 48 : 48–59.

TURNER, W. A., G. CHARTRON, F. LAVILLE, B. MICHELET (1988), Packaging information for peer review: new co-word analysis techniques. In: VAN RAAN, A. F. J (Ed.), *Handbook of Quantitative Studies of Science and Technology*. Netherlands: Elsevier Science Publishers, pp. 291–323.

WEEBER, M., H. KLEIN, A. R. ARONSON, J. G. MORK, L. T. W. DE JONG-VAN DEN BERG, R. VOS (2000), Text-based discovery in biomedicine: the architecture of the DAD-system. In: OVERHAGE, J. M. (Ed.). *Proceedings of the 2000 AMIA Annual Fall Symposium*. Philadelphia, PA: Hanley and Belfus, pp. 903–907.

WEEBER, M., H. KLEIN, L. T. W. DE JONG-VAN DEN BERG, R. VOS (2001), Using concepts in literature-based discovery: Simulating Swanson's Raynaud-fish oil and migraine-magnesium discoveries, *Journal of the American Society for Information Science and Technology*, 52 : 548-557.

Appendix
Details of PubMed searches and cluster analyses

Theme	Search term(s) or phrase(s) (combined by "OR")	Field restriction	Publication window	No. of documents retrieved	No. of unique MeSH / RN terms occuring in at least two documents	Minimal E_{ij}	No. of links	Minimal number of terms per cluster	No. of clusters	Graphically displayed
Raynaud's Disease	RAYNAUD*	TI	1966 - 1985	801	464	0.050	1383	3	44	Figure 1
Fish Oils	FISH OIL FISH OILS COD LIVER OIL CODLIVER OILS SALMON OIL SALMON OILS MENHADEN OIL MENHADEN OILS EICOSAPENTAENOIC ACID EICOSAPENTAENOIC ACIDS	TI	1966 - 1985	248	296	0.050	1753	3	31	Figure 2
Blood Viscosity	BLOOD VISCOSITY	TI	1966 - 1985	500	392	0.050	1286	3	40	Figure 3 a
Blood Viscosity	BLOOD VISCOSITY	MH	1966 - 1985	3645	1694	0.030	3856	2	213	Figure 3 b
Platelet Aggregation	PLATELET AGGREGATION	TI	1966 - 1985	2633	1517	0.050	3796	3	145	Figure 4
Thrombosis	THROMBOSIS	TI	1966 - 1985	7145	2087	0.050	3380	3	194	Figure 5
Migraine	MIGRAINE	TI	1966 - 1987	2581	1089	0.050	2319	3	104	Figure 6
Spreading Depression	SPREADING DEPRESSION SPREADING CORTICAL DEPRESSION	TI	1966 - 1987	327	294	0.050	1246	3	29	Figure 7
Epilepsy	EPILEPSY	TI	1966 - 1987	6681	2155	0.030	3843	3	193	Figure 8
Platelet Aggregation	PLATELET AGGREGATION	TI	1966 - 1987	3020	1665	0.030	3892	3	159	Figure 9
Hypoxia, Brain	HYPOXIA, BRAIN	MH	1966 - 1987	2019	1381	0.050	4287	3	130	Figure 10
Anorexia nervosa	ANOREXIA NERVOSA	MH	1966 - 1987	3131	1625	0.030	4985	3	152	Figure 11
Schizophrenia	SCHIZOPHRENIA	TI	1966 - 1987	7099	2054	0.050	3343	3	179	Figure 12
Prions	PRION PRIONS CREUTZFELDT BOVINE SPONGIFORM KURU SCRAPIE GERSTMANN STRAUSSLER SCHEINKER	TI	1966 - 1995	2848	1341	0.050	3562	3	135	-
Chimeric proteins	CHIMERIC PROTEIN CHIMERIC PROTEINS	ALL	1966 - 1995	2611	2311	0.050	7145	3	217	Figure 13

"*" denotes end truncation.
TI: Title; MH: Medical Subject Heading; ALL: All Fields