Combining Word Semantics within Complex Hilbert Space for Information Retrieval

Peter Wittek[1], Bevan Koopman[2,3], Guido Zuccon[2], and Sándor Darányi[1]

[1] University of Borås, Borås, Sweden
[2] Australian e-Health Research Centre, CSIRO, Brisbane, Australia
[3] Queensland University of Technology, Brisbane, Australia

Abstract. Complex numbers are a fundamental aspect of the mathematical formalism of quantum physics. Quantum-like models developed outside physics often overlooked the role of complex numbers. Specifically, previous models in Information Retrieval (IR) ignored complex numbers. We argue that to advance the use of quantum models of IR, one has to lift the constraint of real-valued representations of the information space, and package more information within the representation by means of complex numbers. As a first attempt, we propose a complex-valued representation for IR, which explicitly uses complex valued Hilbert spaces, and thus where terms, documents and queries are represented as complex-valued vectors. The proposal consists of integrating distributional semantics evidence within the real component of a term vector; whereas, ontological information is encoded in the imaginary component. Our proposal has the merit of lifting the role of complex numbers from a computational byproduct of the model to the very mathematical texture that unifies different levels of semantic information. An empirical instantiation of our proposal is tested in the TREC Medical Record task of retrieving cohorts for clinical studies.

1 Introduction

In quantum theory, states are represented by vectors defined on a complex-valued Hilbert space. Complex numbers are a fundamental aspect in the mathematical formalism of quantum physics. For example, mathematically, the quantum interference term in the law of total probability for disjoint events arises because the probability amplitudes of events are modelled by complex numbers. Quantum-like formalisms were proposed to model systems outside of physics, for example in cognitive science, decision making, economy, etc. In information retrieval (IR), the pioneering work by van Rijsbergen [1] showed that the quantum formalism encompasses many state-of-the-art retrieval models; subsequent works proposed many quantum-like models for IR [2]. Common to all these proposals is the assumption that information objects (queries, documents, etc.) are represented in real-valued Hilbert spaces, even when the key modelling aspect is the quantum interference phenomenon [3]. Zuccon and Piwowarski argued that this assumption is not imposed by the models themselves, which, being grounded on

the mathematics of quantum theory, allow for complex valued representations. Instead it is rooted in the difficulties of understanding how complex numbers could be obtained from term counts in documents [4].

We derive a complex-valued representation of information by encoding semantics by complex numbers. The proposal helps to increase the "semantic power" of traditional semantic space models for IR, by combining different sources of semantic evidence. The intuition underlying our proposal stems from the observation that different models of semantic spaces for IR and text categorisation [5–8] apparently share a joint limitation: they are minimally semantic only, inasmuch as all they utilise is the distributional meaning of words based on their co-occurrence in context. Although not measured explicitly, this limitation probably constrains their performance as well. This long-standing convention has been relying on the use of real numbers only. Our proposal of creating a complex-valued information object grounded in semantic information allows to go beyond the use of mere distributional semantics and includes, e.g., paradigmatic vs. syntagmatic elements of word meaning [9, 10] to merge term co-occurrence statistics with ontological information, e.g., from WordNet [11].

Using an example from medical IR and departing from an earlier study on the use of semantics in such a task [12], we empirically demonstrate how a more sophisticated model of word semantics is implemented in Hilbert space by means of complex numbers. While the retrieval performance is better, we must point out that storage requirements double.

The paper is structured as follows. In Section 2, we briefly discuss the major types of bi- and tripartite theories of word meaning which were consulted to engineer mathematical objects with a higher than usual capacity for information representation. Section 3 inspects random indexing used for model building. Section 4 discusses experiment design. Section 5 interprets the results and Section 6 concludes the paper.

2   From signs to meaning: Engineering sign spaces

Next, we provide a brief account of the theories of meaning that were consulted for engineering the complex-valued representation of word meaning proposed.

In semiotics, a sign (e.g., a word, as well as non-linguistic symbols or clues) is defined as a unity of form and content. Signs are characterised by many important characteristics, which are captured by the following typology:

- Signs are located in some concrete or abstract space vs. having a temporal, e.g., causal nature;
- They are bi- or tripartite. Bipartite theories go back to Aristotle (*form* vs. *substance*) and St. Augustine (*content* vs. *expression*); they found their way into

20th century general linguistics thanks to Ferdinand de Saussure whose sign as word form vs. word meaning was contrasted by American semiotics, itself relying on a tripartite sign concept, adding pragmatics as a third component by the question *for whom* is a sign meaningful;

– Wherever we happen onto a meaningful word or sentence, it is always an instance of a sign, which lends importance to designing such spaces for computer processing.

When focusing on linguistic signs, an alternative way to distinguish between kinds of word meaning is to juxtapose "meaning is use" (i.e., the distributional hypothesis proposed by Wittgenstein [13], Harris [14] and Firth [15]), with "meaning is change" (stimulus-response theories of Bloomfield [16], Morris [17] and Uexkull [18]), and "meaning is equivalence" (referential theories, e.g., by Peirce [19] and Frege [20]). It is of great importance that because none of these theories are able to account for word semantics alone, one has to regard word meaning as being composite. This in turn leads to the insight that unless this compound nature of word meaning is encapsulated in mathematical objects, less progress beyond today's IR models can be realistically expected. The proposal put forward aims to address this observation.

Metric spaces are often used to represent signs by assigning word meaning as substance to a certain location expressed by its coordinates as form. What lends importance to such "charged" locations goes back to another aspect of word semantics specified by Harris' distributional hypothesis [14], stating that words able to replace one another in the same context have highly similar meaning. This is the cornerstone upon which the meaningfulness of semantic spaces rests. Our approach here will be to expand on this practice by unifying language use with its conceptual underpinnings, and merge them as the form and content side of signs, the building blocks of using complex vector space for information representation.

3 Bringing different semantics together: A complex-valued representation of information

Inspired by the observation made in Section 2, we hypothesise that a combination of different sources or types of semantic information is achieved using complex numbers for the representation of information objects in IR models. Such a proposal would provide a means for generating a representation of information based on complex numbers, which could form the basis for more advanced quantum-like models for IR.

In this section, we describe how our proposal is instantiated, i.e., how a complex-valued representation of terms, documents and queries are generated

that brings together different types of word semantics. To do so, we combined previous techniques for random indexing and concept-based document indexing, which are briefly outlined in the next paragraphs.

3.1 Random indexing

Random indexing does not rely on the use of computationally intensive matrix decomposition algorithms like singular value decomposition (SVD) to achieve a fairly low-dimensional representation of a document or term space. This makes random indexing a much more scalable technique in practice as it builds an incremental word space model in a two-step process as follows [7, 21]:

- First, every context (e.g., each document or each word) in the data is assigned a unique and randomly generated representation called an index vector. These index vectors are sparse, high-dimensional, and ternary, that is, their dimensionality (k) is in the order of thousands, and they consist of a small number of randomly distributed +1s and -1s, with the rest of the elements of the vectors set to 0;
- Then, context vectors are produced by scanning through the text, and every time a word occurs in a context (e.g., in a document, or within a sliding window), that context's k-dimensional index vector is added to the context vector for the word in question. Words are thus represented by k-dimensional context vectors that are effectively the sum of the words' contexts.

Document vectors are simply the sums of their constituent word vectors, hence the document space is also k-dimensional. The number of dimensions is defined by k, and random indexing does not provide an explicit way of computing it, being a parameter of the model. These dimensions are not topics, in contrast to other low-dimensional embeddings such as latent semantic indexing [5]. Efficient and extendable open source implementations of random indexing exist [22].

3.2 Concept-based document indexing

In concept-based indexing, documents are represented by concepts rather than terms, as is instead the case for traditional term-based representations. Concepts are usually defined by an ontology or are knowledge-based, and different strings of text are represented by the same concept, indicating that these have identical meaning. For example, in the medical domain, the expressions "heart attack" and "myocardial infarction" have the same meaning and are usually mapped to the same underlying concept. The simpler form of concept-based indexing consists of extracting concepts from the textual content of documents and then

| Unique terms | 218,574 | Unique concepts | 36,467 |
| Total terms | 40,212,729 | Total concepts | 67,183,177 |

Table 1. Term and concept statistics for the TREC Medical Records Track collection.

representing documents as a bag-of-concepts (BOC) vector, as opposed to the traditional bag-of-words (BOW) approach. More advanced forms of concept-based indexing have been proposed; for example, in concept-based indexing for medical IR, Koopman et al., 2012, and Zuccon et al., 2012 capture the relations (implicit or formal, respectively) between concepts encoded in the ontology of reference [23, 24]. These approaches are, however, beyond our scope.

We consider a simple BOC representation, where documents correspond to vectors of concept identifiers, and we thus assume that a mapping between strings of texts and concepts exist. In the experiments of Section 4, we use the procedure outlined in Koopman et al. [12], which involves converting both queries and documents to concepts[4] using the medical natural language processing system called MetaMap [26]. Both documents and queries are thus represented not as their original terms but as concept identifiers from the SNOMED-CT ontology. A standard IR indexing and retrieval process can then be applied to the concept documents and queries.

A concept may correspond to an n-gram of text, e.g., concept 165664003 refers to "Entire articular process of cervical vertebra"; likewise, an n-gram (or just a single word) may stand for several concepts. As such, the term statistical behaviour observed in language does not always apply to concept representations. In particular, concept representations does not obey Zipf's law, i.e., do not exhibit the typical long tail distribution of very infrequent concepts that instead a term representation commonly exhibits [27]. Similarly, the number of concepts used to represent a corpus of documents differ greatly from the size of the term vocabulary for that same corpus. To exemplify this, we anticipate the statistics obtained from the representations of the document corpus used in the experiments of Section 4. Table 1 summarises the term and concept statistics obtained when indexing the TREC Medical Records Track collection; more details about the indexing procedure are given later. The table shows that while the corpus contains many more concepts than terms, the vocabulary size for concepts is one order of magnitude smaller than that for terms.

3.3 Combining word semantics: Documents in complex space

Our proposal revolves around using complex numbers to combine different forms of semantic information. In the following, we consider two instances of semantic information, namely distributional and referential. Specifically, we

---

[4] Specifically, concepts from the SNOMED-CT medical ontology [25].

draw distributional semantic information using the random indexing technique described in Section 3.1; whereas, referential semantic information is drawn from the concept-based indexing procedure outlined in Section 3.2. The concept index vectors were assigned to SNOMED-CT concepts, and corresponding document vectors were derived from these by superposition. The position of an index term with a bipartite sign nature in the complex vector space is composed from term frequency and other statistics in a term-document matrix, and represented as the real component of the resulting complex vector, with the representation of concepts from the concept-based indexing constituting its imaginary component. Thereby in any complex term weight, the real component encoded distributional semantics whereas the imaginary component hosted referential semantics. Such weights are then used to build complex term, document and query vectors for retrieval.

Because of the difference in vocabulary sizes between the term and concept representations, the dimensionality of the term space and that of the concept space does not match. The use of random indexing provides a solution to address this issue, where the dimension of the random vectors is used to force a common dimensionality among the two sources of information, as detailed in the following. Assume a document-term matrix of dimension $m \times n$ is built from the corpus of documents, where m is the number of documents and n is the size of the term vocabulary. Similarly, assume that the corresponding document-concept matrix for that corpus is of dimension $m \times p$, with p being the number of concepts in the concept-vocabulary. By applying random indexing to both matrices maintaining the number of dimensions of the random vectors to the same k, not only distributional semantic information is extracted from the respective original matrices, but compatible representations of the two spaces are also obtained. That is, the random indexing representation for of the original document-term matrix will have a dimensionality of $m \times k$; similarly, the random indexing representation for concepts will have a dimensionality of $m \times k$, eliminating issues of dimensionality mismatches between vectors from term or concept representations. In the representation proposed here, document similarity calculated as the inner product between complex-valued vectors reflects both the distributional and the ontological facets of document content. Similarly, the comparison of the vectors associated to the real part of the representation with those associated with the imaginary part would provide the similarity between the statistical term space representation of a document and its ontological concept space representation.

Merging the two in a complex space is a trivial exercise, and it allows for measurements of phase between a statistical term space vs. an ontological concept space (see Figure 1), with the inner product of document similarity reflecting both the distributional and the ontological facets of content.
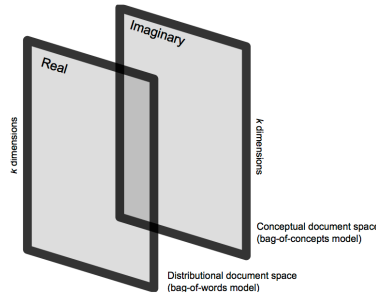
Fig. 1. Schematic representation of a complex document space.

The method proposed here is similar to that suggested by van Rijsbergen, where different functions of term statistics (specifically, term frequency and inverse document frequency) are assigned to the real and imaginary components of complex numbers [1]. That approach provided poor retrieval effectiveness, and the power of representation was questioned [4]. However, differently from van Rijsbergen's proposal, we suggest to encode in the components of the complex numbers semantic information (purely distributional information and higher-level conceptual information) rather than functions of term statistics. Another Hilbert space-based representation embedded the two kinds of semantics by a seriation of the feature space and using the L_2 space of square-integrable functions [28]. Compared to that approach, the only preprocessing we need is mapping the terms to concepts, which is faster than seriation.

To aid the understanding of our method, we present a simple example. Consider the following documents: D_1 = "`kidney stones`", D_2 = "`kidney`" and D_3 = "`renal calculi`", and the query Q = "`kidney stones`". A term-based retrieval system processing Q would return only the documents D_1, D_2 (in this order). However, `renal calculi` in D_3 is actually a synonym of the query `kidney stones`, while D_2 is actually not relevant. Therefore, D_3 should be ranked higher than D_2. Using our method, when a concept representation is included, the phrases `kidney stones` and `renal calculi` both map to the same SNOMED CT concept `155868000`. Thus, our ranking approach would retrieve D_1 in the first rank position because of the contribution of both term and concept weighting; and D_3 above D_2 because of the inclusion of the concept weighting.

4   Empirical investigation

Next, we outline the experiment we devised to test our proposal. It describes an initial effort to evaluate the merit of the complex-value representation; further validation will be subject of future work.

4.1 Experiment settings

To benchmark the efficiency of the proposed representation, we evaluated the method in the IR task provided in the TREC Medical Records Track, which consists in retrieving medical records of patients that satisfy clinical and demographic criteria specified as queries. We followed previous work in this area for combining medical records belonging to a single patient into a unique document, called a patient visit document [23, 24]. We followed the procedure outlined by Koopman et al. [12] to obtain a concept representation. Consistently with previous work, SNOMED-CT was chosen as ontology of reference for the concept representation. In total, the collection consisted of 17,198 patient visit documents and 81 queries; statistics for both term and concept representations are outlined in Table 1. Retrieval effectiveness was measured by mean average precision (MAP) and precision of the top 10 ranked documents (P@10), as well as a precision-recall analysis.

Measure	Term-based	Concept-based	Complex
MAP	0.0886	0.1084	0.1245
P@10	0.1593	0.1963	0.2235

Table 2. MAP and P@10 for term retrieval, concept retrieval, and our method (labeled "complex" retrieval).
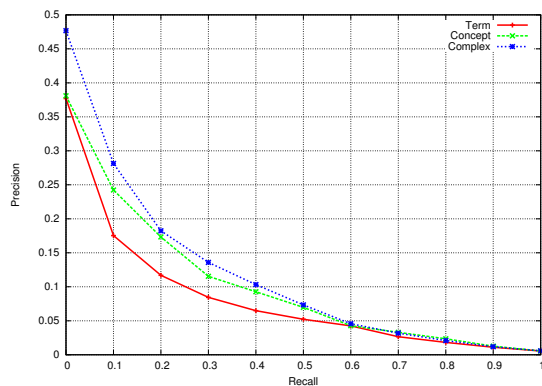

Fig. 2. 11 points precision-recall graph.

We used Lucene 3.6.2 to index and retrieve documents, while the SemanticVectors package, version 3.8 [22, 29], was used to construct a random indexing representation for terms and concepts bag of words. The context window of a word or a concept was the full document. The use of the respective random indexes for retrieval formed the two baselines, we contrasted our method against in this evaluation, namely term-based (BOW) random indexing and concept-based (BOC) random indexing. To implement our method, we constructed a complex space by using the random indices of the BOW and BOC models as described in

Section 3.3. Queries were represented in the complex space in the same way as the documents. The number k of random dimensions was set to 200 for all methods; experimentation with other dimensions is left for future work, but initial results showed effectiveness not to vary considerably when changing k.

For the standard BOW and BOC models, the distance function used to judge the similarity between documents and queries was the inner product of the corresponding real space (as formed by the random index). The counterpart of the inner product in the complex space is the Hermitian product, but this yields a complex number. Since complex numbers do not have a natural ordering, this product cannot be used for ranking. We adapted the inner product to measure the overlap between the real and complex part in the same way as in a real space, yielding a real number. This approach is different from the similarity measure defined for complex spaces in SemanticVectors. None of the spaces were normalised, as the inner product is insensitive to the norm of vectors.

4.2   Results

The retrieval effectiveness of the methods is reported in Table 2. The mean average precision of the method that exploits the complex-valued representation is found to be 15% higher than that of the concept-based approach, and 40% higher compared to the term-based random indexing approach. Similar findings are obtained when considering P@10. The analysis of the 11-point precision-recall interpolation across all queries is showed in Figure 2. The results suggest that precision is markedly higher at lower recall levels; that is, the proposed complex-valued representation retrieves more relevant documents in the top results than baselines methods considered here.

The results obtained by all methods considered here are generally lower than those reported by state-of-the-art IR method on the same task [30]. This suggests that random indexing alone is not an effective method for document retrieval in the medical domain. Similar findings were obtained when using random indexing for query expansion in this task [31], although that method delivered higher effectiveness than the results presented here. Nevertheless, in our evaluation, we are interested in understanding the value the complex-valued representation adds to the baseline methods, rather than the actual absolute effectiveness of the instantiation investigated here. More effective instantiations of our proposal may in fact consider distributional semantic techniques other than semantic indexing, boosting effectiveness.

We further analysed the empirical results obtained in our experiments by examining the values of the angles formed by the real and imaginary components of the complex vectors representing documents. Figure 3 shows the distribution
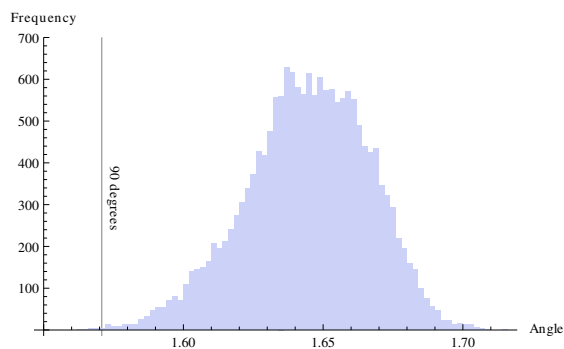
Fig. 3. Distribution of angles between real and imaginary components of complex vectors representing documents from the TREC Medical Records Track collection. Angles were expressed in radians and were normalised in the range $[0, 2\pi[$.

of these angles in radians normalised in the range $[0, 2\pi[$, while Table 3 reports the mean, maximum and minimum values of the (unnormalised) angles.

Mean absolute phase	Lowest Phase	Highest Phase
6.9520	-39.9163	28.6791

Table 3. Mean value and extrema of phase

We observed that documents for which the angles assumed the lowest phase value in radians across the collection (< -30) generally contained a large amount of numerical laboratory test results, as well as long lists of medication brand names and dosages. This sort of content is often not mapped to concepts by MetaMap, which led to large amounts of information found in the term representation missing in the concept representation.

Vice versa, we observed that documents characterised by large angles ($>$ 20 radians) are often short in length (due to the small number of terms), but when mapped to concepts, their concept representation was considerably longer than the term-based one (i.e., a document is represented by more concepts than terms). This is the case when MetaMap assigns multiple concepts to the same (or overlapping) strings of text. This finding suggests that in such cases, the concept representation provides a more accurate description of the content of short documents than the original term representation.

5 Conclusions

Complex numbers have a key role in the mathematical framework of quantum theory; however they have been overlooked when using quantum-like formalisms for modelling systems outside of physics: this is the case especially in IR. We departed from previous approaches that considered complex numbers as a simple

computational byproduct of the models, and we proposed a novel representation of information based on complex numbers which brings together different sources of semantics. Specifically, in the empirical instantiation of our proposal, we mixed distributional (contextual) and referential (ontological) semantics, although the proposed approach is not limited to the particular techniques used to derive semantics. The empirical evaluation of the proposed complex-valued representation has lead to encouraging results, although further evaluation is required. We must point out that the storage requirement complex-valued representation the sum of the a term-based and concept-based model, which means it is nearly double than the space required by either. Given the sparse data structures, this trade-off is hardly an issue.

The proposed model allows for the measurement of quantitative distances between the distributional form of a word or a phrase and its referential representation in a concept space. Technically this distance corresponds to the angle between the real and imaginary components of information objects (terms, documents, etc.). The representation invites the study of the compositionality of meaning in noun compounds from a new perspective, a field of intensive study within Quantum Interaction [32, 33]. In addition, user relevance is an indirect but inherent component in our framework as the sources of referential semantics within the representation (e.g., SNOMED-CT in the medical domain) are constructed with domain-specific relevance in mind. A query effectively interacts with both the distributional pattern of terms and the underlying concept space where relevance judgements are implicitly encoded.

References

1. van Rijsbergen, C.J.: The Geometry of Information Retrieval. Cambridge University Press, New York, NY, USA (2004)
2. Song, D., Lalmas, M., van Rijsbergen, K., Frommholz, I., Piwowarski, B., Wang, J., Zhang, P., Zuccon, G., Bruza, P.D., Arafat, S., et al.: How quantum theory is developing the field of information retrieval. In: Proc. of QI, Arlington, VA, USA (November 2010) 105–108
3. Zuccon, G., Azzopardi, L.: Using the quantum probability ranking principle to rank interdependent documents. In: Proc. of ICTIR, Cambridge, UK (September 2009) 357–369
4. Zuccon, G., Piwowarski, B., Azzopardi, L.: On the use of complex numbers in quantum models for information retrieval. In: Proc. of ICTIR, Bertinoro, Italy (September 2011)
5. Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Harshman, R.: Indexing by latent semantic analysis. JASIST **41**(6) (1990) 391–407
6. Lund, K., Burgess, C.: Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior Research Methods Instruments and Computers **28** (1996) 203–208
7. Kanerva, P., Kristofersson, J., Holst, A.: Random indexing of text samples for latent semantic analysis. In: Proc. of CogSci. Volume 1036., Philadelphia, PA, USA (2000)
8. Karlgren, J., Sahlgren, M.: From words to understanding. In Uesaka, Y., Kanerva, P., Asoh, H., eds.: Foundations of Real-World Intelligence. CSLI Publications (2001) 294–308

9. Sahlgren, M.: The Word-Space Model: using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. PhD thesis, Institutionen för lingvistik. Department of Linguistics, Stockholm University (2006)

10. Symonds, M., Bruza, P., Sitbon, L., Turner, I.: Modelling word meaning using efficient tensor representations. In: Proc. of PacLic. (November 2011) 313–322

11. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA, USA (1998)

12. Koopman, B., Bruza, P., Sitbon, L., Lawley, M.: Towards semantic search and inference in electronic medical records: an approach using concept-based information retrieval. AMJ **9** (2012) 482–488

13. Wittgenstein, L.: Philosophical Investigations. Blackwell Publishing, Oxford, UK (1967)

14. Harris, Z.: Distributional structure. In Harris, Z., ed.: Papers in structural and transformational Linguistics. Formal Linguistics. Humanities Press, New York, NY, USA (1970) 775–794

15. Firth, J.R.: Papers in Linguistics 1934—1951. London: Oxford University Press (1957)

16. Bloomfield, L.: Language. Holt, Reinhart and Winston, New York, NY, USA (1933)

17. Morris Charles, W.: Signs, Language and Behavior. New York: Prentice Hall (1946)

18. von Uexküll, J.: The theory of meaning. Semiotica **42**(1) (1982) 25–82

19. Peirce, C.: Logic as semiotic: The theory of signs. In Peirce, C., Buchler, J., eds.: Philosophical Writings of Peirce. Dover Publications (1955) 98–119

20. Frege, G.: Sense and reference. The Philosophical Review **57**(3) (1948) 209–230

21. Sahlgren, M.: An introduction to random indexing. In: Proc. of TKE, Copenhagen, Denmark (August 2005)

22. Widdows, D., Ferraro, K.: Semantic vectors: a scalable open source package and online technology management application. In: Proc. LREC, Marrakech, Morocco (May 2008)

23. Koopman, B., Zuccon, G., Bruza, P., Sitbon, L., Lawley, M.: Graph-based concept weighting for medical information retrieval. In: Proc. of ADCS, Dunedin, New Zealand (December 2012) 80–87

24. Zuccon, G., Koopman, B., Nguyen, A., Vickers, D., Butt, L.: Exploiting medical hierarchies for concept-based information retrieval. In: Proc. of ADCS, Dunedin, New Zealand (December 2012) 111–114

25. Spackman, K.: SNOMED Clinical Terms Basics. International Health Terminology Standards Development Organisation Technical Report (August 2008)

26. Aronson, A.R., Lang, F.M.: An overview of MetaMap: historical perspective and recent advances. JAMIA **17**(3) (2010) 229–236

27. Wu, S.T., Liu, H., Li, D., Tao, C., Musen, M.A., Chute, C.G., Shah, N.H.: Unified medical language system term occurrences in clinical notes: a large-scale corpus analysis. JAMIA **19**(e1) (2012) e149–e156

28. Wittek, P., Tan, C.L.: Compactly supported basis functions as support vector kernels for classification. IEEE Trans. Pattern Anal. Mach. Intell. **33**(10) (2011) 2039 –2050

29. Widdows, D., Cohen, T., De Vine, L.: Real, complex, and binary semantic vectors. In: Proc. of QI, Paris, France, Springer (June 2012) 24–35

30. Voorhees, E., Tong, R.: Overview of the TREC Medical Records Track. In: Proc. of TREC, Gaithersburg, MD, USA (November 2011)

31. Wu, S., Masanz, J., Ravikumar, K., Liu, H.: Three questions about clinical information retrieval. In: Proc. of TREC, Gaithersburg, MD, USA (November 2012)

32. Aerts, D., Czachor, M.: Quantum aspects of semantic analysis and symbolic artificial intelligence. Journal of Physics A: Mathematical and General **37** (2004) L123–L132

33. Bruza, P., Kitto, K., Ramm, B., Sitbon, L., Song, D., Blomberg, S.: Quantum-like non-separability of concept combinations, emergent associates and abduction. Logic Journal of IGPL **20**(2) (2012) 445–457