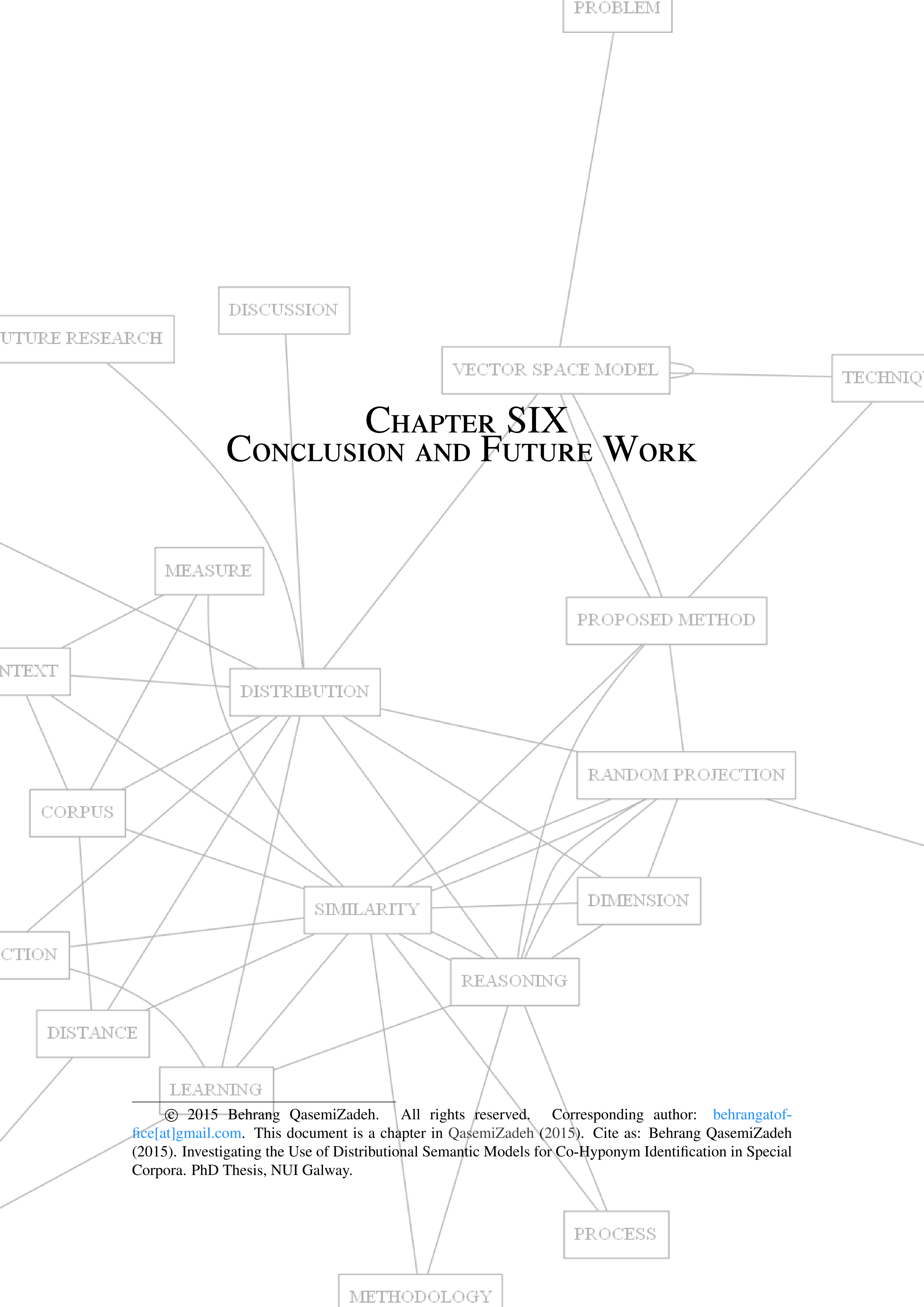

Chapter SIX
Conclusion and Future Work

This page is intentionally left blank.

Contents

This page is intentionally left blank.

Chapter 6

Conclusion and Future Work

This thesis began with an emphasis on the importance of the extraction of co-hyponym terms—that is, terms that characterise a particular category of concepts in a knowledge domain—for facilitating the process of knowledge acquisition from text. It is explained how the principles of distributional semantics and automatic term extraction can be combined to bridge the semantic gap, to decipher the meaning of terms, and to address the task of extracting co-hyponym terms. Using random projections, vector space models with reduced dimensionality are constructed to represent the distributional properties of terms. In turn, an example-based learning framework is utilised to implement a similarity-based reasoning mechanism in order to identify co-hyponym terms. This thesis details the design and evaluation of the proposed methodology.

This chapter is divided into three sections. Section 6.1 restates the research contributions. Section 6.2 discusses a number of open research questions, as well as a few topics for future research. Finally, Section 6.3 concludes this thesis by a short summary.

6.1 Research Contributions

6.1.1 The Proposed Method for Identifying Co-Hyponym Terms

The main contribution of this thesis is a new perspective of and a novel approach to the identification of co-hyponym terms: a problem that has been so far overlooked in the acquisition of knowledge from text. The proposed distributional representation of terms in a vector space model and the use of similarity-based reasoning for deciphering their meaning (see Section 5.1 and 5.2, Chapter 5) alleviate a number of concerns that arise with respect to the flexibility and the user-friendliness of previously proposed techniques.

Categorisation in general, and, in particular term categorisation, is a major mechanism for organising knowledge and improving the performance of information systems. Based on the specification of a proposed abstraction from a knowledge domain, concepts (and thus terms) are organised in a taxonomy consisting of several co-hyponym groups—each group containing a number of terms that share a *type-of* relationship with a common general concept (i.e., a hypernym). The fluid nature of knowledge is inevitably reflected by the changes in the way that the domain knowledge is abstracted, and, in turn, how these co-hyponym terms are defined in a knowledge structure.

To maintain, and to embody this dynamic structure using tightly supervised techniques—such as those employed in the development of entity extraction systems—is labour-intensive, and thus, expensive. In contrast to these methods, the approach proposed in this thesis is flexible and easy to maintain. In the proposed approach, the mechanism employed for the representation of terms' semantics (i.e., the vector space representation) is independent of the devised categorisation of terms. An update in the structure of knowledge is applied by providing new examples of a newly emerged taxon. Neither text annotation nor a training process is required for the development of a model (meta-language) that captures the structure of terms. In addition, the suggested incremental technique for the construction of a vector space model allows the model to be updated at any time during its use—for example, new terms can be added and removed, and the vectors that represent them can be updated independently of each other. Similarly, examples that are employed by the similarity-based reasoning framework can be updated (new examples are added or removed, and the existing ones modified) at any time during the model's life cycle.

From the perspective of a user of such a system—perhaps, an expert in the knowledge domain, who may have minimal or no training in natural language processing—the process of adapting an existing extraction system to a new task, domain, or even a new class of terms is cumbersome. If a rule-based methodology is employed, new rules must be devised; if a supervised learning technique is employed, a new model must be developed. In contrast, the proposed methodology is user-friendly and intuitive in the sense that the user requires to provide only a few samples of what is, and perhaps what is not, desirable in a new category of terms. Coupled with the system's flexibility in the manipulation of vectors, feedback from the user can be easily incorporated into the system during its life cycle.

Lastly, the proposed technique is scalable both vertically and horizontally. The fixed dimension of vectors, which can be set and known prior to the extraction task, allows one to implement the method for an effective exploitation of the computational resources available to the system in a single node. This is particularly advantageous if GPU-accelerated[1] computing techniques are employed for similarity measurements. Needless to say, the combination of random projections and example-based similarity reasoning exploited in the proposed method suits parallel, distributed computing (e.g., using the *MapReduce* programming paradigm) extremely well.

Additional novel characteristics of the method proposed for identifying co-hyponym terms are listed in Section 1.2.

6.1.2 A Systematic Evaluation of the Proposed Method

A systematic evaluation of the method proposed in this thesis is carried out (see Section 5.4 of Chapter 5). In the experiments performed (see Sections 5.3), the viability of the proposed distributional hypothesis for identifying co-hyponym terms is verified.[2]

Several parameters that play a role in the performance of the proposed method and the reciprocal relationship between these parameters are investigated (see Discussion in Section 5.5). The discussion of the experiments in Chapter 5 focusses on finding the best configuration of the parameters for the context-window (i.e., the way co-occurrence frequencies are collected) and the parameters for the example-based learning method (i.e., similarity-based reasoning). The interdependence among these parameters, including the figure of merit employed for assessing the performance of the method (i.e., precision at small recall vs. large recall values), is also an important consideration. This is confirmed by the method for reporting observations.

In the experiments described here, context-windows are configured differently with respect to their size (i.e., the extent to which they stretch in the vicinity of candidate terms), direction (i.e., left, right, or around the candidate terms), and encoding information about the order of the words they contain. The example-based learning framework is evaluated with respect to the measure employed for computing similarities (i.e., the cosine measure, the Euclidean distance, and the city block distance) as well as the role of neighbourhood size selection (i.e., the number of examples taken into consideration during the weighting procedure). Moreover, the performance of the method is examined under different conditions, namely in the absence and the presence of noise, for corpora of different sizes, for several categories of concepts, and at both small and large recall values. Despite a number of similarities in the results, no single best configuration could be recommended for all the tasks (see Section 5.5). However, the following settings for the parameters can be recommended:

- With respect to the size of the context-window (shown by t):
  - often $2 \leq t \leq 4$ is sufficient. However, if the corpus is small or the targeted co-hyponym terms are infrequent, then a large t such as $4 \leq t \leq 6$ could be a better choice. This is particularly so if the set of candidate terms contains many invalid

[1] That is, graphical processing units that are customised for linear algebra calculations.

[2] See also RQ 1 to 4.

terms and small recall values are intended. At the same time, choosing a large value of t can introduce noise and thus decrease the performance, particularly if a distance metric is employed for computing similarities.

- With respect to the direction in which the context-window is extended to collect co-occurrences:

  – if the size of the corpus and the intended recall value are small, and the set of candidate terms contains many invalid terms, then context-windows that extend around candidate terms are a better choice. Otherwise, context-windows that expand to the left of candidate terms to collect co-occurrences with their preceding words are recommended.

- With respect to information about the sequential order of words in context-windows:

  – encoding this information does not necessarily enhance performance. If the corpus is large (or, the targeted co-hyponym terms are frequent), and the context-window is extended to the left of candidate terms, encoding the word order information can enhance the result by as much as 10%. It is observed that encoding this information often improves the performance of the best performing models whereas it diminishes the performance of other models.

- With respect to the selection of a similarity measure:

  – a distance measure is, perhaps, a better choice if a small recall value is intended or the intended recall is small in relation to the number of reference terms. However, the cosine measure is an obligate choice if a large recall value is intended, or the intended recall is very large in relation to the number of reference terms. Similarly, if the set of candidate terms contains a large number of invalid terms that share a common context with valid terms,[1] then cosine is a better choice than a distance metric.

- With respect to the neighbourhood size (k) selection in the k-nearest neighbours framework:

  – if a small recall value is intended and cosine is employed for similarity measurements, then the nearest neighbour can outperform other choices of k. Otherwise, a large value of k is recommended. Particularly, if a distance metric is chosen, a large k is a more reliable choice than a small one. In addition, when the corpus becomes larger, a smaller value of k can be employed.

[1] For instance, as suggested in Chapter 5, when valid terms appear nested in invalid terms (e.g., the appearance of the valid term *computational linguistics* in invalid candidate terms such as *in computational linguistic studies*, *interesting computational linguistics*, and so on).

6.1.3 The Method for Incremental Construction of Vector Spaces

In this thesis, novel techniques for the incremental construction of vector spaces, particularly ℓ_1-normed spaces, are introduced (see Chapter 4). The proposed methods are employed to obviate the *curse of dimensionality* problem. The mathematical theorems behind the previously employed technique, known as the random indexing (RI) method, is explained and ameliorated by a guideline for setting its parameters. It is shown that RI is an incremental method for the construction of ℓ_2-normed spaces (i.e, Euclidean spaces), which is based on the principle of sparse random projections.[1]

The aforementioned principles are employed to introduce the random Manhattan indexing technique (RMI) and a variation of it named random Manhattan integer indexing (RMII). Both RMI and RMII implement random projections in ℓ_1-normed spaces using projections of randomly created matrices with an asymptotic *Cauchy* distribution. However, by a slight alteration in the distribution of the random projection matrices and a new distance estimator, the RMII method avoids floating point arithmetics during the construction of a vector space. This thesis employs proposed incremental vector space construction techniques for identifying co-hyponym terms. These, however, can be also used in many text analyses algorithms that employ vector space mathematics in general, and in big text data analytics in particular.

Finally, by the help of the principles that are introduced to justify RI, RMI, and RMII, a mathematical justification of a method known as the permutation technique is provided (see Section 5.3.2.3.1, Chapter 5). The permutation technique is employed to capture and to encode into a vector space model the sequential order of words in a text. In this thesis, the previous intuitive justification of the permutation method is complemented using the newly provided mathematical discussion in Chapter 4.

6.2 Open Questions and Future Work

From a very broad perspective, getting machines to understand natural languages, as they are used by people to communicate with and to understand each other, has been, and perhaps, will be one of the biggest research challenges for curious minds. What is obvious is that language as *an instrument of communication*[2] is a non-random complex system. However, distinguishing useful patterns in this system, and translating them into machine-accessible semantics has remained an open research question.[3] In addition to this kind of question, the research presented in this thesis can be extended in several ways, as described below.

[1] See also RQ 5.

[2] Note that this is not necessarily an exclusive function of language, but one of many.

[3] For example, will it be possible to find a comprehensive representation of text data other than the text data itself that meets all the requirements for a system with natural language understanding capability?

6.2.1 Semantic Compositionality

In the context of distributional semantics, the *compositionality* of semantics and meaning is currently receiving much attention (see Mitchell and Lapata, 2010, for an overview). Apart from numerous research efforts (e.g., see Baroni et al., 2014; Coecke et al., 2011), many debates are also going on with respect to the limits and the theoretical foundations of the *compositionality* of semantics (e.g., see Goldberg, 2015). In compositional distributional semantics, research is focused on inferring the meanings of a linguistic entity from its smaller parts—such as words from morphemes, and phrases and sentences from words—using an algebraic structure (e.g., the vector space model studied and employed in this thesis).

For instance, in a number of approaches, given a vector space model of word co-occurrences and a finite number of mathematical operations such as adding, subtracting, and so on, the goal is to answer whether it is possible to infer the meaning of a multiword expression from the vectors of the words that construct the expression (e.g., see Kiela and Clark, 2013). Evidently, this research overlaps with the study of the meanings of terms, particularly *complex*[1] terms. The semantic compositionality of terms is not dealt with in the research presented here: *Are terms, particularly complex terms, irreducible linguistic units such as idioms?* Or, *do they show a degree of compositionality?*

A systematic study of the aforementioned question is one way to extend the proposed research in this thesis (e.g., by limiting the scope of the research proposed in Baldwin et al., 2003, to terminology). As discussed in Chapter 5, complex terms are very rare in special corpora; as a result, the collected co-occurrences in special corpora show a very long-tail statistical distribution.[2] If terms have compositional semantics, then the proposed techniques in compositional distributional semantics can be also used to address problems arising from a lack of data for collecting evidence that is required for establishing the meaning of terms.

6.2.2 Term Space Models for Relations Other Than Co-Hyponymy

Term space models implemented in this thesis are employed and evaluated for identifying co-hyponymy relationships between terms. However, these models can be used to recognise relationships between terms other than co-hyponymy—for example, synonymy,[3] associative and relatedness relationships, etc. This is similar to the applications of these methods in general language lexicography, which has been recently encouraged for terminology, too (e.g., see Faber and L'Homme, 2014).[4] If co-hyponymy relationships between terms are employed to suggest an organisation of a specialised vocabulary, then identifying *compatible* and *incompatible* co-hyponyms seems an interesting future research.[5]

As briefly suggested in Chapter 1, the problem of *is-a* overload can be expected in this

[1] That is, multi-token.

[2] Longer than the distribution of the co-occurrences of words in general language corpora.

[3] That is, to address the term variation problem.

[4] See also Chapter 3.

[5] For example, to find *disjoint* classes in a domain ontology.

context. Investigating methods address this problem is also an interesting future research. The advantages of similarity-based reasoning offered by the term space methodology can be used as a complementary mechanism, not only to extract useful information from text but also to facilitate logical inference mechanisms. There is an exciting potential for integrating existing (*semi-*)manually-built formal knowledge resources (e.g., the open schemas and data contributed by the semantic Web research community) and distributional semantic models to build a comprehensive system of reasoning (e.g., see Angeli and Manning, 2014). To make this potential a reality, terminology—as a research discipline—could be the point of convergence for the systematic integration of these research efforts. That is to say, the suggested perspective in terminology[1] can provide a coherent theoretical basis for rational integration of empiricist corpus-based distributional methods and rationalist formal knowledge representation frameworks.

6.2.3 Extending the Scope of Evaluation

Extending evaluated context parameters and enhancing performance

This thesis evaluated the performance of the proposed method using the so-called *flat* distributional models—that is, no linguistic information, such as part-of-speech categories, lemmatisation, or syntactic relationships are employed during the construction of the models. Whereas constructing a flat model demands low computational power and scales out easily, the use of linguistic information could enhance the performance.[2] The evaluation presented here can thus be extended by taking into account the linguistic properties of context elements (i.e., the co-occurred words with candidate terms).

Moreover, context-windows are configured only for a few parameters. This can be easily extended. In the evaluations performed in this thesis, those context-windows that extend around terms are assumed to be *symmetrical* (e.g., 5 tokens to the left and 5 tokens to the right side of terms: that is, 5+5). However, these context-windows can be extended *asymmetrically* (e.g., 5 tokens to the left and 1 tokens to the right side of terms: that is, 5+1). The influence of extending context-windows asymmetrically in the performance of the method can be thus studied in the future. Exclusion of words in context-windows is also a possibility that can be investigated, too. For example, context-windows do not require to be extended in the immediate vicinity of terms, but with an offset of a few tokens (e.g., as suggested by Brown et al., 1992). This method for defining context-windows can perhaps reduce noise resulted from errors in identifying candidate terms.

Likewise, the evaluation can be extended by using various weighting mechanisms other than the raw frequencies of words, and similarity measures other than cosine and the ℓ_2 and the ℓ_1 distances. In this study, the evaluation is limited to the use of a fixed set of reference vectors. Investigating methods for choosing the best representative reference vectors would be another way to extend the reported evaluation. Although this question has been investigated from the data analytics perspective (e.g., see Garcia et al., 2012), it is interesting to explore linguistic characteristics of such instances.

[1] Which goes beyond the interpretation of *terms as labels for concepts*; see Chapter 3.

[2] See also related discussion in Chapter 2.

As suggested in Chapter 5, bootstrap learning is a plausible solution for improving the method's performance when large recall values are intended; this is one of the limits of the method. In this case, a number of new parameters are introduced. For example, the way the set of reference vectors is extended and the way concept drifting is controlled. This must be investigated together with other parameters of the method.

Evaluation across sublanguages and domains

The evaluation presented in this thesis is limited to the scientific sublanguage from of the molecular biology domain (i.e., the GENIA corpus). Although our initial observations in a sublanguage other than molecular biology (see Zadeh and Handschuh, 2014b,c) is similar to the reported results here, further empirical investigations can be helpful to have a better understanding of the method's behaviour across sublanguages and to further demonstrate its applicability across domains.

Qualitative study of the method's output

The presented quantitative evaluation can be complemented by a qualitative evaluation.[1] The method's parameters for instance can be investigated with respect to the various characteristics of terms they extract. For example, Weeds et al. (2004) study the frequency characteristics of extracted words using different similarity measures. A similar approach can be adopted for studying the method's parameters and the effect of these parameters on various aspects of the properties of the extracted terms (e.g., the frequency of terms, their generality-specificity, etc.).

Not presented in the reported evaluations is the identification of co-hyponyms in *nested* and *hierarchical* structures. For instance, the category of protein terms in the GENIA corpus is made of several sub-categories. The fine-grained identification of these categories of concepts and their evaluation can be beneficial for a number of tasks. The extracted set of co-hyponym terms using the proposed distributional model often consists of entries that are synonyms, metonyms, and hypernyms (as suggested in the previous sentence). Identifying these entries can enhance the quality of the generated set of co-hyponym terms (e.g., similar to what is addressed by Weeds et al., 2014, for words in a general vocabulary).

Modelling additional elements of the communicative context

Last but not least, extending the evaluation parameters to additional elements of the communicative context is another interesting research quest. For example, extending a distributional model to learn from user interactions and integrating a model of behaviour in the underlying distributional model[2] is an interesting research with many practical applications. Whereas current research is focused mostly on the learning algorithms, the distributional semantic framework allows for flexible expression of this type of information in the knowledge base itself, instead of the learning (training) mechanism.

[1] See the discussion on the evaluation of term extraction methods in Section 3.7 of Chapter 3.

[2] Other than, or in addition to, the manipulation of the set of reference terms (as is implied in Section 5.6 of chap. 5), such as the proposed solution for automatic spell checking in QasemiZadeh et al. (2006).

Diachronic investigation

In the presented study, the evaluation is limited to the extraction of co-hyponym terms at a synchronic level. However, a diachronic analysis of term categories (as well as their meanings), which has a number of important applications, such as *trend analysis*, remains an open research area. The investigation of diachronic aspects of terminology in particular, and in general adding a temporal dimension to distributional semantic models, is certainly an exciting untouched research challenge. The lack of systematic studies in such an important area is, perhaps, due to the lack of suitable language resources.

As reported in Zadeh and Handschuh (2014a), we are developing a language resource, named *ACL RD-TEC*, that can be used for investigating diachronic aspects of a terminology. The ACL RD-TEC dataset consists of manually annotated terms from scientific publications that are drawn from the ACL anthology reference corpus (ACL ARC). The ACL ARC is a fixed set of 10,921 scientific publications in the domain of computational linguistics from 1965 to 2006 (Bird et al., 2008). Term annotations in ACL RD-TEC can thus be mapped to this time-line in order to provide a benchmark for diachronic study of terms and their meanings.

Investigating interaction with the domain conceptualisation

The conceptualisation of a domain defines the co-hyponym groups, which the method proposed in this thesis identifies. This conceptualisation is dynamic and varies even from one person to another, as discussed in Chapter 1. The conceived granularity of concepts is especially important in the performance of the method (not only from the statistical point of view, but also from the linguistic and knowledge engineering perspectives). The presented evaluation does not answer questions that arise with respect to this factor. The design of an evaluation framework that can assess this interaction is thus necessary (e.g., as suggested by Rindflesch and Fiszman, 2003).

6.2.4 Further Generalisation of Random Projections

Random projections are modern mathematical tools, which are still relatively unexplored, both theoretically and empirically. This thesis proposed a new incremental technique for constructing vector space models using random projections. The discussion about these projections is limited to α-normed spaces, where $\alpha = 1$, or 2. However, as suggested in Chapter 4, the proposed methodology can be extended to α-normed spaces other than $\alpha = 1$, or 2. The application of these random projections in distributional semantics for the construction of vector space models remains an untouched research avenue. Whether these techniques are suitable for various text analytic applications, however, is an open research question that must be addressed in future research and through experiments.

In this thesis, a single random projection is employed for the construction of vector spaces. However, it is possible to combine random projections in different normed spaces and in different ways. For example, instead of using a single random projection from an n-dimensional to an m-dimensional space of which $n \ll m$, one can apply two different random projections; a projection from the n-dimensional space to an m_1-dimensional space, and then from the m_1-dimensional space to the m-dimensional space of which

$n \ll m_1 \ll m$.[1] Using this multi-stage projection allows the approximation of similarities to be carried out in different normed space, if desirable. In addition, a trade-off between the dimension of the projected spaces and the expected errors in the approximated similarities can be considered,[2] allowing for a more efficient computation of similarities and perhaps enhancing the time complexity of a similarity-based reasoning process over big text data—a similar rational as is employed in *locality-sensitive hashing* techniques and *space partitioning* (e.g., see Datar et al., 2004; Dhesi and Kar, 2010).

6.3 Summary

To summarise, this thesis aimed at designing a framework for characterising the conceptual organisation of terms in a specialised vocabulary induced from a domain-specific corpus. To meet this goal, the construction of distributional semantic models with fixed reduced dimensionality using random projection techniques is studied. With the help of a similarity-based reasoning mechanism, the application of these models to characterise co-hyponymy relationships between terms is investigated.

[1] Note that the trending multi-layer neural networks (i.e., the so-called *deep learning* techniques) are also based on the same mathematical principle.

[2] Since $m_1 \ll m$, it is expected that the approximated distances in the m_1-dimensional space are more accurate than the m-dimensional space.

Reference List

Angeli, G. and Manning, C. D. (2014). NaturalLI: Natural logic inference for common sense reasoning. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 534–545, Doha, Qatar. Association for Computational Linguistics. 203

Baldwin, T., Bannard, C., Tanaka, T., and Widdows, D. (2003). An empirical model of multiword expression decomposability. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 89–96, Sapporo, Japan. Association for Computational Linguistics. 202

Baroni, M., Bernardi, R., and Zamparelli, R. (2014). Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies*, 9(6):5–110. 202

Bird, S., Dale, R., Dorr, B., Gibson, B., Joseph, M., Kan, M.-Y., Lee, D., Powley, B., Radev, D., and Tan, Y. F. (2008). The ACL anthology reference corpus: A reference dataset for bibliographic research in computational linguistics. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., and Tapias, D., editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, pages 1755–1759, Marrakech, Morocco. European Language Resources Association. 205

Brown, P. F., de Souza, P. V., Mercer, R. L., Pietra, V. J. D., and Lai, J. C. (1992). Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–479. 203

Coecke, B., Sadrzadeh, M., and Clark, S. (2011). Mathematical foundations for a compositional distributional model of meaning. In van Benthem, J. and Moortgat, M., editors, *Linguistic Analysis*, volume 36(1–4) of *A Festschrift for Joachim Lambek*, pages 345–384. 202

Datar, M., Immorlica, N., Indyk, P., and Mirrokni, V. S. (2004). Locality-sensitive hashing scheme based on p-stable distributions. In *SCG '04: Proceedings of the Twentieth Annual Symposium on Computational Geometry*, pages 253–262, Brooklyn, NY, USA. ACM. 206

Dhesi, A. and Kar, P. (2010). Random projection trees revisited. In Lafferty, J. D., Williams, C. K. I., J., S.-T., Zemel, R. S., and Culotta, A., editors, *Advances in Neural Information Processing Systems 23*. Curran Associates, Inc. 206

Faber, P. and L'Homme, M.-C. (2014). Lexical semantic approaches to terminology: An introduction. *Terminology*, 20(2):143–150. 202

Garcia, S., Derrac, J., Cano, J. R., and Herrera, F. (2012). Prototype selection for nearest neighbor classification: Taxonomy and empirical study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):417–435. 203

Goldberg, A. E. (2015). Compositionality. In Riemer, N., editor, *The Routledge Handbook of Semantics*. Route ledge. 202

Kiela, D. and Clark, S. (2013). Detecting compositionality of multi-word expressions using nearest neighbours in vector space models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1427–1432, Seattle, Washington, USA. Association for Computational Linguistics. 202

Mitchell, J. and Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388–1429. 202

QasemiZadeh, B. (2015). *Investigating the Use of Distributional Semantic Models for Co-Hyponym Identification in Special Corpora*. PhD thesis, National University of Ireland, Galway. i

QasemiZadeh, B., Ilkhani, A., and Ganjeii, A. (2006). Adaptive language independent spell checking using intelligent traverse on a tree. In *2006 IEEE Conference on Cybernetics and Intelligent Systems*, pages 1–6, Bangkok, Thailand. IEEE. 204

Rindflesch, T. C. and Fiszman, M. (2003). The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *Journal of Biomedical Informatics*, 36(6):462–477. Unified Medical Language System. 205

Weeds, J., Clarke, D., Reffin, J., Weir, D., and Keller, B. (2014). Learning to distinguish hypernyms and co-hyponyms. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2249–2259, Dublin, Ireland. Dublin City University and Association for Computational Linguistics. 204

Weeds, J., Weir, D., and McCarthy, D. (2004). Characterising measures of lexical distributional similarity. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland. Association for Computational Linguistics. 204

Zadeh, B. Q. and Handschuh, S. (2014a). The ACL RD-TEC: A dataset for benchmarking terminology extraction and classification in computational linguistics. In Drouin, P., Grabar, N., Hamon, T., and Kageura, K., editors, *COLING 2014: Computerm 2014: 4th International Workshop on Computational Terminology: Proceedings of the Workshop*, pages 52–63, Dublin, Ireland. Association for Computational Linguistics and Dublin City University. 205

Zadeh, B. Q. and Handschuh, S. (2014b). Evaluation of technology term recognition with random indexing. In Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, Reykjavik, Iceland. European Language Resources Association. ACL Anthology Identifier: L14-1703. 204

Zadeh, B. Q. and Handschuh, S. (2014c). Investigating context parameters in technology term recognition. In Meyers, A., He, Y., and Grishman, R., editors, *Proceedings of the COLING Workshop on Synchronic and Diachronic Approaches to Analyzing Technical Language (SADAATL 2014)*, pages 1–10. Association for Computational Linguistics and Dublin City University. 204

This page is intentionally left blank.