Every child should have parents: a taxonomy refinement algorithm based on hyperbolic term embeddings

Rami Aly[1], **Shantanu Acharya**[2], **Alexander Ossa**[1], **Arne Köhn**[4,1],
Chris Biemann[1], and **Alexander Panchenko**[3,1]

[1]Universität Hamburg, Hamburg, Germany
[2]National Institute of Technology Mizoram, Aizawl, India
[3]Skolkovo Institute of Science and Technology, Moscow, Russia
[4]Saarland University, Saarbrücken, Germany
{5aly,2ossa,koehn,biemann,panchenko}@informatik.uni-hamburg.de

Abstract

We introduce the use of Poincaré embeddings to improve existing state-of-the-art approaches to domain-specific taxonomy induction from text as a signal for both relocating wrong hyponym terms within a (pre-induced) taxonomy as well as for attaching disconnected terms in a taxonomy. This method substantially improves previous state-of-the-art results on the SemEval-2016 Task 13 on taxonomy extraction. We demonstrate the superiority of Poincaré embeddings over distributional semantic representations, supporting the hypothesis that they can better capture hierarchical lexical-semantic relationships than embeddings in the Euclidean space.

1 Introduction

The task of taxonomy induction aims at creating a semantic hierarchy of entities by using hyponym-hypernym relations – called *taxonomy* – from text corpora. Compared to many other domains of natural language processing that make use of pre-trained dense representations, state-of-the-art taxonomy learning is still highly relying on traditional approaches like extraction of lexical-syntactic patterns (Hearst, 1992) or co-occurrence information (Grefenstette, 2015). Despite the success of pattern-based approaches, most taxonomy induction systems suffer from a significant number of disconnected terms, since the extracted relationships are too specific to cover most words (Wang et al., 2017; Bordea et al., 2016). The use of distributional semantics for hypernym identification and relation representation has thus received increasing attention (Shwartz et al., 2016). However, Levy et al. (2015) observe that many proposed supervised approaches instead learn prototypical hypernyms (that are hypernyms to many other terms), not taking into account the relation between both terms in classification. Therefore, past applications of distributional semantics appear to be rather unsuitable to be directly applied to taxonomy induction as the sole signal (Tan et al., 2015; Pocostales, 2016). We address that issue by introducing a series of simple and parameter-free refinement steps that employ word embeddings in order to improve existing domain-specific taxonomies, induced from text using traditional approaches in an unsupervised fashion.

We compare two types of dense vector embeddings: the standard word2vec CBOW model (Mikolov et al., 2013a,b), that embeds terms in *Euclidean space* based on distributional similarity, and the more recent Poincaré embeddings (Nickel and Kiela, 2017), which capture similarity as well as hierarchical relationships in a *hyperbolic space*. The source code has been published[1] to recreate the employed embedding, to refine taxonomies as well as to enable further research of Poincaré embeddings for other semantic tasks.

2 Related Work

The extraction of taxonomic relationships from text corpora is a long-standing problem in ontology learning, see Biemann (2005) for an earlier survey. Wang et al. (2017) discuss recent advancements in taxonomy construction from text corpora. Conclusions from the survey include: i) The performance of extraction of IS-A relation can be improved by studying how pattern-based and distributional approaches complement each other; ii) there is only limited success of pure deep learn-

---

[1]https://github.com/uhh-lt/Taxonomy_Refinement_Embeddings

ing paradigms here, mostly because it is difficult to design a single objective function for this task.

On the two recent TExEval tasks at SemEval for taxonomy extraction (Bordea et al., 2015, 2016), attracting a total of 10 participating teams, attempts to primarily use a distributional representation failed. This might seem counterintuitive, as taxonomies are surely modeling semantics and thus their extraction should benefit from semantic representations. The 2015 winner INRIASAC (Grefenstette, 2015) performed relation discovery using substring inclusion, lexical-syntactic patterns and co-occurrence information based on sentences and documents from Wikipedia. The winner in 2016, TAXI (Panchenko et al., 2016), harvests hypernyms with substring inclusion and Hearst-style lexical-syntactic patterns (Hearst, 1992) from domain-specific texts obtained via focused web crawling. The only submission to the TExEval 2016 task that relied exclusively on distributional semantics to induce hypernyms by adding a vector offset to the corresponding hyponym (Pocostales, 2016) achieved only modest results. A more refined approach to applying distributional semantics by Zhang et al. (2018) generates a hierarchical clustering of terms with each node consisting of several terms. They find concepts that should stay in the same cluster using embedding similarity – whereas, similar to the TExEval task, we are interested in making distinctions between all terms. Finally, Le et al. (2019) also explore using Poincaré embeddings for taxonomy induction, evaluating their method on hypernymy detection and reconstructing WordNet. However, in contrast to our approach that filters and attaches terms, they perform inference.

3 Taxonomy Refinement using Hyperbolic Word Embeddings

We employ embeddings using distributional semantics (i.e. word2vec CBOW) and Poincaré embeddings (Nickel and Kiela, 2017) to alleviate the largest error classes in taxonomy extraction: the existence of *orphans* – disconnected nodes that have an overall connectivity degree of zero and *outliers* – a child node that is assigned to a wrong parent. The rare case in which multiple parents can be assigned to a node has been ignored in the proposed refinement system. The first step consists of creating domain-specific Poincaré embeddings (§ 3.1). They are then used to identify and

relocate outlier terms in the taxonomy (§ 3.2), as well as to attach unconnected terms to the taxonomy (§ 3.3). In the last step, we further optimize the taxonomy by employing the endocentric nature of hyponyms (§ 3.4). See Figure 1 for a schematic visualization of the refinement pipeline. In our experiments, we use the output of three different systems. The refinement method is generically applicable to (noisy) taxonomies, yielding an improved taxonomy extraction system overall.

3.1 Domain-specific Poincaré Embedding

Training Dataset Construction To create domain-specific Poincaré embeddings, we use noisy hypernym relationships extracted from a combination of general and domain-specific corpora. For the general domain, we extracted 59.2 GB of text from English Wikipedia, Gigaword (Parker et al., 2009), ukWac (Ferraresi et al., 2008) and LCC news corpora (Goldhahn et al., 2012). The domain-specific corpora consist of web pages, selected by using a combination of BootCat (Baroni and Bernardini, 2004) and focused crawling (Remus and Biemann, 2016). Noisy IS-A relations are extracted with lexical-syntactic patterns from all corpora by applying PattaMaika[2], PatternSim (Panchenko et al., 2012), and WebISA (Seitner et al., 2016), following (Panchenko et al., 2016).[3]

The extracted noisy relationships of the common and domain-specific corpora are further processed separately and combined afterward. To limit the number of terms and relationships, we restrict the IS-A relationships on pairs for which both entities are part of the taxonomy's vocabulary. Relations with a frequency of less than three are removed to filter noise. Besides further removing every reflexive relationship, only the more frequent pair of a symmetric relationship is kept. Hence, the set of cleaned relationships is transformed into being antisymmetric and irreflexive. The same procedure is applied to relationships extracted from the general-domain corpus with a frequency cut-off of five. They are then used to expand the set of relationships created from the domain-specific corpora.

---

[2] http://jobimtext.org: The PattaMaika component is based on UIMA RUTA (Kluegl et al., 2016).

[3] Alternatively to the relations extracted using lexical patterns, we also tried to use hypernyms extracted using the pretrained HypeNet model (Shwartz et al., 2016), but the overall taxonomy evaluation results were lower than the standard baseline of the TAXI system and thus are not presented here.
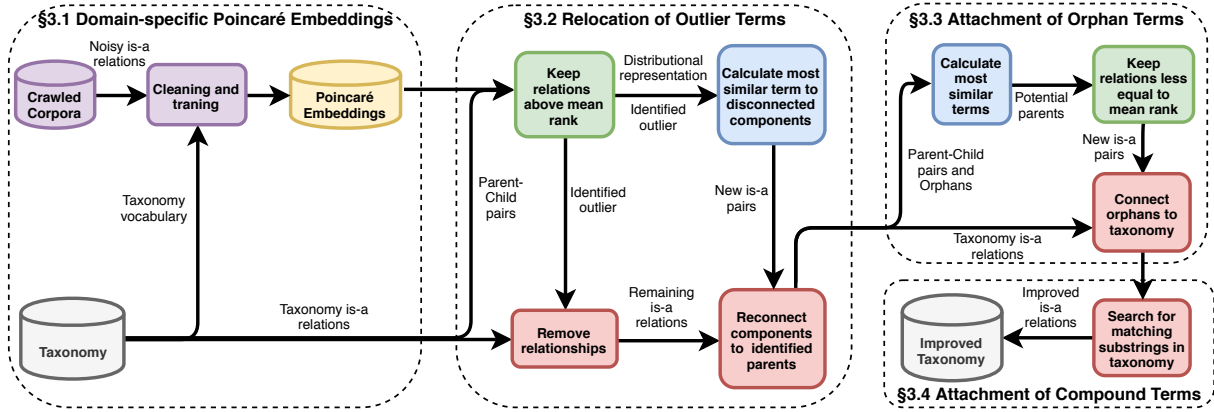
Figure 1: Outline of our taxonomy refinement method, with paper sections indicated.

Hypernym-Hyponym Distance Poincaré embeddings are trained on these cleaned IS-A relationships. For comparison, we also trained a model on noun pairs extracted from WordNet (P-WN). Pairs were only kept if both nouns were present in the vocabulary of the taxonomy. Finally, we trained the word2vec embeddings, connecting compound terms in the training corpus (Wikipedia) by '_' to learn representations for compound terms, i.e multiword units, for the input vocabulary.

In contrast to embeddings in the Euclidean space where the cosine similarity $\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}$ is commonly applied as a similarity measure, Poincaré embeddings use a hyperbolic space, specifically the Poincaré ball model (Stillwell, 1996). Hyperbolic embeddings are designed for modeling hierarchical relationships between words as they explicitly capture the hierarchy between words in the embedding space and are therefore a natural fit for inducing taxonomies. They were also successfully applied to hierarchical relations in image classification tasks (Khrulkov et al., 2019). The distance between two points $\mathbf{u}, \mathbf{v} \in \mathcal{B}^d$ for a d-dimensional Poincaré Ball model is defined as:

$$d(\mathbf{u}, \mathbf{v}) = \text{arcosh}\Big(1 + 2\frac{||\mathbf{u} - \mathbf{v}||^2}{(1 - ||\mathbf{u}||^2)(1 - ||\mathbf{v}||^2)}\Big).$$

This Poincaré distance enables us to capture the hierarchy and similarity between words simultaneously. It increases exponentially with the depth of the hierarchy. So while the distance of a leaf node to most other nodes in the hierarchy is very high, nodes on abstract levels, such as the root, have a comparably small distance to all nodes in the hierarchy. The word2vec embeddings have no notion of hierarchy and hierarchical relationships cannot be represented with vector offsets across

the vocabulary (Fu et al., 2014). When applying word2vec, we use the observation that distributionally similar words are often co-hyponyms (Heylen et al., 2008; Weeds et al., 2014).

3.2 Relocation of Outlier Terms

Poincaré embeddings are used to compute and store a rank $rank(x, y)$ between every child and parent of the existing taxonomy, defined as the index of y in the list of sorted Poincaré distances of all entities of the taxonomy to x. Hypernym-hyponym relationships with a rank larger than the mean of all ranks are removed, chosen on the basis of tests on the 2015 TExEval data (Bordea et al., 2015). Disconnected components that have children are re-connected to the most similar parent in the taxonomy or to the taxonomy root if no distributed representation exists. Previously or now disconnected isolated nodes are subject to orphan attachment (§ 3.3).

Since distributional similarity does not capture parent-child relations, the relationships are not registered as parent-child but as co-hyponym relationships. Thus, we compute the distance to the closest co-hyponym (child of the same parent) for every node. This filtering technique is then applied to identify and relocate outliers.

3.3 Attachment of Orphan Terms

We then attach orphans (nodes unattached in the input or due to the removal of relationships in the previous step) by computing the rank between every orphan and the most similar node in the taxonomy. This node is an orphan's potential parent. Only hypernym-hyponym relationships with a rank lower or equal to the mean of all stored ranks are added to the taxonomy. For the word2vec system, a link is added between the parent of the most

similar co-hyponym and the orphan.

3.4 Attachment of Compound Terms

In case a representation for a compound noun term does not exist, we connect it to a term that is a substring of the compound. If no such term exists, the noun remains disconnected. Finally, the Tarjan algorithm (Tarjan, 1972) is applied to ensure that the refined taxonomy is asymmetric: In case a circle is detected, one of its links is removed at random.
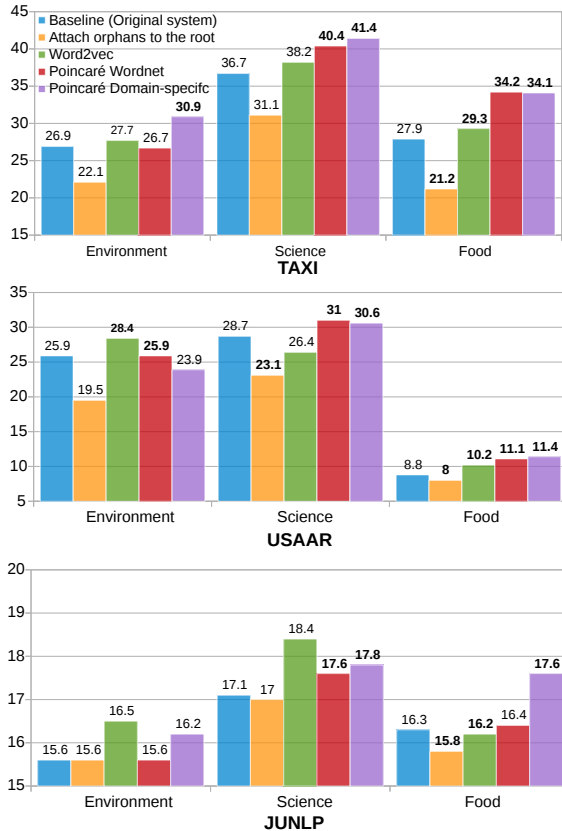
Figure 2: F_1 results for the systems on all domains. Vocabulary sizes: environment ($|V| = 261$), science ($|V| = 453$), food ($|V| = 1556$). **Bold** numbers are significantly different results to the original system with $p < 0.05$.

4 Evaluation

Proposed methods are evaluated on the data of SemEval2016 TExEval (Bordea et al., 2016) for submitted systems that created taxonomies for all domains of the task[4], namely the task-winning system TAXI (Panchenko et al., 2016) as well as the systems USAAR (Tan et al., 2016) and JUNLP (Maitra and Das, 2016). TAXI harvests

---

[4] http://alt.qcri.org/semeval2016/task13/index.php

hypernyms with substring inclusion and lexical-syntactic patterns by obtaining domain-specific texts via focused web crawling. USAAR and JUNLP heavily rely on rule-based approaches. While USAAR exploits the endocentric nature of hyponyms, JUNLP combines two string inclusion heuristics with semantic relations from BabelNet. We use the taxonomies created by these systems as our baseline and additionally ensured that taxonomies do neither have circles nor in-going edges to the taxonomy root by applying the Tarjan algorithm (Tarjan, 1972), removing a random link from detected cycles. This causes slight differences between the baseline results in Figure 2 and (Bordea et al., 2016).

5 Results and Discussion

Comparison to Baselines Figure 2 shows comparative results for all datasets and measures for every system. The *Root* method, which connects all orphans to the root of the taxonomy, has the highest connectivity but falls behind in scores significantly. Word2vec CBOW embeddings partly increase the scores, however, the effect appears to be inconsistent. Word2vec embeddings connect more orphans to the taxonomy (cf. Table 2), albeit with mixed quality, thus the interpretation of word similarity as co-hyponymy does not seem to be appropriate. Word2vec as a means to detect hypernyms has shown to be rather unsuitable (Levy et al., 2015). Even more advanced methods such as the *diff* model (Fu et al., 2014) merely learn so-called *prototypical hypernyms*.

Both Poincaré embeddings variants outperform the word2vec ones yielding major improvements over the baseline taxonomy. Employing the McNemar (1947) significance test shows that Poincaré embeddings' improvements to the original systems are indeed significant. The achieved improvements are larger on the TAXI system than on the other two systems. We attribute to the differences of these approaches: The rule-based approaches relying on string inclusion as carried out by USAAR and JUNLP are highly similar to step §3.4. Additionally, JUNLP creates taxonomies with many but very noisy relationships, therefore step §3.3 does not yield significant gains, since there are much fewer orphans available to connect to the taxonomy. This problem also affects the USAAR system for the food domain. For the environment domain, however, USAAR creates a

Word	Parent in TAXI	Parent after refinement	Gold parent	Closest neighbors
second language acquisition	—	linguistics	linguistics	applied linguistics, semantics, linguistics
botany	—	genetics	plant science, ecology	genetics, evolutionary ecology, animal science
sweet potatoes	—	vegetables	vegetables	vegetables, side dishes, fruit
wastewater	water	waste	waste	marine pollution, waste, pollutant
water	waste, natural resources	natural resources	aquatic environment	continental shelf, management of resources
international relations	sociology, analysis, humanities	humanities	political science	economics, economic theory, geography

Table 1: Example words with respective parent(s) in the input taxonomy and after refinement using our domain-specfic Poincaré embeddings, as well as the word's closest three neighbors (incl. orphans) in embeddings.

Domain	word2vec	P. WordNet	P. domain-specific	# orphans
Environment	25	18	34	113
Science	56	39	48	158
Food	347	181	267	775

Table 2: Number of attached orphans in taxonomies created by TAXI using different embeddings.

taxonomy with very high precision but low recall which makes step §3.2 relatively ineffective. As step §3.3 has shown to improve scores more than §3.2, the gains on JUNLP are comparably lower.

WordNet-based Embeddings The domain-specific Poincaré embeddings mostly perform either comparably or outperform the WordNet-based ones. In error analysis, we found that while WordNet-based embeddings are more accurate, they have a lower coverage as seen in Table 2, especially for attaching complex multiword orphan vocabulary entries that are not contained in WordNet, e.g., *second language acquisition*. Based on the results we achieved by using domain-specific Poincaré embeddings, we hypothesize that their attributes result in a system that learns hierarchical relations between a pair of terms. The closest neighbors of terms in the embedding clearly tend to be more generic as exemplarily shown in Table 1, which further supports our claim. Their use also enables the correction of false relations created by string inclusion heuristics as seen with *wastewater*. However, we also notice that few and inaccurate relations for some words results in imprecise word representations such as for *botany*.

Multilingual Results Applying domain-specific Poincaré embeddings to other languages also creates overall improved taxonomies, however the scores vary as seen in Table 3. While the score of all food taxonomies increased substantially, the taxonomies quality for environment did not improve, it even declines. This seems to be due to the lack of extracted relations in (§3.1), which results in imprecise representations and a highly limited

vocabulary in the Poincaré embedding model, especially for Italian and Dutch. In these cases, the refinement is mostly defined by step §3.4.

Language	Domain	Original	Refined	# rel. data	# rel. gold
English	Environment	26.9	**30.9**	657	261
	Science	36.7	**41.4**	451	465
	Food	27.9	**34.1**	1898	1587
French	Environment	23.7	**28.3**	114	266
	Science	31.8	**33.1**	118	451
	Food	22.4	**28.9**	598	1441
Italian	Environment	**31.0**	30.8	2	266
	Science	32.0	**34.2**	4	444
	Food	16.9	**18.5**	57	1304
Dutch	Environment	**28.4**	27.1	7	267
	Science	29.8	**30.5**	15	449
	Food	19.4	**21.8**	61	1446

Table 3: F_1 comparison between original (TAXI) and refined taxonomy using domain-specific embeddings.

6 Conclusion

We presented a refinement method for improving existing taxonomies through the use of hyperbolic Poincaré embeddings. They consistently yield improvements over strong baselines and in comparison to word2vec as a representative for distributional vectors in the Euclidean space. We further showed that Poincaré embeddings can be efficiently created for a specific domain from crawled text without the need for an existing database such as WordNet. This observation confirms the theoretical capability of Poincaré embeddings to learn hierarchical relations, which enables their future use in a wide range of semantic tasks. A prominent direction for future work is using the hyperbolic embeddings as the sole signal for taxonomy extraction. Since distributional and hyperbolic embeddings cover different relations between terms, it may be interesting to combine them.

Acknowledgments

References

Marco Baroni and Silvia Bernardini. 2004. Bootcat: Bootstrapping corpora and terms from the web. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, pages 1313–1316, Lisbon, Portugal.

Chris Biemann. 2005. Ontology learning from text: A survey of methods. *LDV Forum*, 20(2):75–93.

Georgeta Bordea, Paul Buitelaar, Stefano Faralli, and Roberto Navigli. 2015. Semeval-2015 task 17: Taxonomy Extraction Evaluation (TExEval). In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 902–910, Denver, CO, USA.

Georgeta Bordea, Els Lefever, and Paul Buitelaar. 2016. SemEval-2016 Task 13: Taxonomy Extraction Evaluation (TExEval-2). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1081–1091, San Diego, CA, USA.

Adriano Ferraresi, Eros Zanchetta, Marco Baroni, and Silvia Bernardini. 2008. Introducing and evaluating ukWaC, a very large web-derived corpus of English. In *Proceedings of the 4th Web as Corpus Workshop. Can we beat Google?*, pages 47–54, Marrakech, Morocco.

Ruiji Fu, Jiang Guo, Bing Qin, Wanxiang Che, Haifeng Wang, and Ting Liu. 2014. Learning semantic hierarchies via word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 1199–1209, Baltimore, MD, USA.

Dirk Goldhahn, Thomas Eckart, and Uwe Quasthoff. 2012. Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In *Proceedings of the Eight International Conference on Language Resources and Evaluation*, pages 759–765, Istanbul, Turkey.

Gregory Grefenstette. 2015. INRIASAC: Simple Hypernym Extraction Methods. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 911–914, Denver, CO, USA.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th Conference on Computational Linguistics*, pages 539–545, Nantes, France.

Kris Heylen, Yves Peirsman, Dirk Geeraerts, and Dirk Speelman. 2008. Modelling word similarity: an evaluation of automatic synonymy extraction algorithms. In *Proceedings of the sixth international language resources and evaluation*, pages 3243–3249, Marrakech, Morocco.

Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. 2019. Hyperbolic image embeddings. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Peter Kluegl, Martin Toepfer, Philip-Daniel Beck, Georg Fette, and Frank Puppe. 2016. UIMA Ruta: Rapid development of rule-based information extraction applications. *Natural Language Engineering*, 22(1):1–40.

Matt Le, Stephen Roller, Laetitia Papaxanthos, Douwe Kiela, and Maximilian Nickel. 2019. Inferring concept hierarchies from text corpora via hyperbolic embeddings. *arXiv preprint arXiv:1902.00913*.

Omer Levy, Steffen Remus, Chris Biemann, and Ido Dagan. 2015. Do supervised distributional methods really learn lexical inference relations? In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 970–976, Denver, CO, USA.

Promita Maitra and Dipankar Das. 2016. JUNLP at SemEval-2016 task 13: A language independent approach for hypernym identification. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1310–1314, San Diego, CA, USA.

Quinn McNemar. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157.

Tomas Mikolov, Kai Chen, G.S Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In *ICLR Workshop*, Scottsdale, AZ, USA.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, Stateline, NV, USA.

Maximillian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems 30*, pages 6338–6347, Long Beach, CA, USA.

Alexander Panchenko, Stefano Faralli, Eugen Ruppert, Steffen Remus, Hubert Naets, Cédrick Fairon, Simone P. Ponzetto, and Chris Biemann. 2016. TAXI at SemEval-2016 Task 13: a taxonomy Induction Method based on Lexico-syntactic Patterns, Substrings and Focused Crawling. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1320–1327, San Diego, CA, USA.

Alexander Panchenko, Olga Morozova, and Hubert Naets. 2012. A semantic similarity measure based on lexico-syntactic patterns. In *Proceedings of KONVENS 2012*, pages 174–178, Vienna, Austria.

Robert Parker, David Graff, Junbo Kong, Ke Chen, and Kazuaki Maeda. 2009. English gigaword forth edition. In *Linguistic Data Consortium*, Philadelphia, PA, USA.

Joel Pocostales. 2016. NUIG-UNLP at SemEval-2016 Task 13: A Simple Word Embedding-based Approach for Taxonomy Extraction. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1298–1302, San Diego, CA, USA.

Steffen Remus and Chris Biemann. 2016. Domain-specific corpus expansion with focused webcrawling. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 3607–3611, Portorož, Slovenia.

Julian Seitner, Christian Bizer, Kai Eckert, Stefano Faralli, Robert Meusel und Heiko Paulheim, and Simone P. Ponzetto. 2016. A large database of hypernymy relations extracted from the Web. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 360–367, Portorož, Slovenia.

Vered Shwartz, Yoav Goldberg, and Ido Dagan. 2016. Improving hypernymy detection with an integrated path-based and distributional method. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 2389–2398, Berlin, Germany.

John Stillwell. 1996. *Sources of hyperbolic geometry*. History of Mathematics, Volume 10. American Mathematical Society.

Liling Tan, Francis Bond, and Josef van Genabith. 2016. USAAR at SemEval-2016 task 13: Hyponym endocentricity. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1303–1309, San Diego, CA, USA.

Liling Tan, Rohit Gupta, and Josef van Genabith. 2015. USAAR-WLV: Hypernym generation with Deep Neural Nets. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 932–937, Denver, CO, USA.

Robert Tarjan. 1972. Depth first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146 –160.

Chengyu Wang, Xiaofeng He, and Aoying Zhou. 2017. A short survey on taxonomy learning from text corpora: Issues, resources and recent advances. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1190–1203, Copenhagen, Denmark.

Julie Weeds, Daoud Clarke, Jeremy Reffin, David Weir, and Bill Keller. 2014. Learning to distinguish hypernyms and co-hyponyms. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*, pages 2249–2259, Dublin, Ireland.

Chao Zhang, Fangbo Tao, Xiusi Chen, Jiaming Shen, Meng Jiang, Brian Sadler, Michelle Vanni, and Jiawei Han. 2018. TaxoGen: Unsupervised topic taxonomy construction by adaptive term embedding and clustering. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2701–2709, London, United Kingdom.