Co-Clustering Triples from Open Information Extraction

Koninika Pal
kpal@mpi-inf.mpg.de
Max Planck Institute for Informatics
Saarbrücken, Germany

Vinh Thinh Ho
hvthinh@mpi-inf.mpg.de
Max Planck Institute for Informatics
Saarbrücken, Germany

Gerhard Weikum
weikum@mpi-inf.mpg.de
Max Planck Institute for Informatics
Saarbrücken, Germany

ABSTRACT

Similar facts are often expressed in different ways in natural language text, which introduces the redundancy and ambiguity of Subject-Predicate-Object (SPO) triples in Open Information Extraction (Open IE). This work focuses on canonicalizing such SPO triples. We propose a clustering framework using non-negative matrix tri-factorization that jointly clusters predicate phrases and subject-object pairs, and aligns them in a meaningful manner. The evaluation shows that our co-clustering method outperforms significantly over rule mining and Knowledge-Base-embedding approaches for two existing datasets.

KEYWORDS

Matrix Factorization, Knowledge Bases, Co-clustering

ACM Reference Format:
Koninika Pal, Vinh Thinh Ho, and Gerhard Weikum. 2020. Co-Clustering Triples from Open Information Extraction. In *7th ACM IKDD CoDS and 25th COMAD (CoDS COMAD 2020), January 5–7, 2020, Hyderabad, India.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3371158.3371183

1 INTRODUCTION

Large Knowledge Bases (KBs), like YAGO [21], DBpedia [1], Freebase [4], Wikidata, contain hundreds of millions of facts about millions of entities. These facts are an important building block to learn, understand, and augment information. They are represented in the form of (ideally canonicalized) triples of (Subject, Predicate, Object) in KBs, e.g., *"Delhi is the capital of India"* can be represented as `(Delhi, capital-of, India)`. Knowledge bases can be constructed using Open IE methods that extract subject, predicate, and object from any noun or verb phrases appearing all over the web. However, such extraction methods lead to the retrieval of redundant facts, as the same information can be expressed differently in natural language text, e.g., the phrase `"birthplace"` can also be presented as `"place of birth"`. Moreover, depending on the context, a verb might convey different information. For example, `"chase"` or `"attack"` can refer to very different meanings: predators chasing and attacking their prey – `(lion, attack, zebra)`, or students chasing a deadline and attacking a problem – `(student, chase, problem)`. Another challenge in Open IE is the retrieval of redundant diverse statements with different granularity labels that share sufficient semantic similarity, e.g., `(panda, are, endangered)` and `(panda, are, endangered in China)`, the latter presents a refinement of the former.

One of the major drawbacks in Open IE is the imbalance among extracted predicates due to the presence of overly generic verb phrases in the language. For example, top-5 predicates cover 32% of TupleKB [12] facts, which are based on very general verb phrases – `"have"`, `"is a"`, `"include"`, `"has part"`, and `"is part of"`. On the other hand, a long tail of rare predicates (600 out of total 1600) occur less than 10 times in this dataset. Recent work on building commonsense KBs [18] shows similar characteristics, where top-5 frequent predicates cover approximately 40% of facts. Both generic and rare predicates become important issues in handling data. The specific semantic relationship between subject and object may lose due to the use of generic predicates, whereas it is difficult to capture statistically significant patterns in the facts with rare predicates.

The prior works of canonicalizing open KBs [7, 17] generally cluster entities and predicates separately, using hierarchical agglomerative clustering (HAC) with different similarity functions. A semi-canonicalized rule mining approach [7] clusters predicates using transitive property of equivalent predicates, where entities must be linked to an existing KB. An ILP-based clustering method [12] is used in TupleKB to canonicalize predicates. However, all of these approaches impose hard clustering, i.e., a phrase belongs to exactly one cluster. Nonetheless, in practice, a predicate can carry different meanings as mentioned earlier with the example of `"chase"` and `"attack"`. Analogously, subject and object arguments can also have ambiguous surface forms that would map to different word senses. Using word sense disambiguation (see [14] for a survey), WebChild [22] addresses to solve this problem, but the additional complexity of the method leads to a high level of noise.

Therefore, our goal is to use a more relaxed *soft clustering* of predicates and subject-object pairs to allow a phrase to be part of multiple clusters. Additionally, to tackle the problem of interpreting predicate phrases, depending on the context of associated subjects and objects, we attempt *co-clustering* method that jointly clusters predicates and subject-object pairs. With such *soft-co-clustering* approach, we aim to achieve the following goals:

(1) *Normalization*: capturing semantically similar predicates such as `"possess"`, `"control"` and `"occupy"`.
(2) *Specialization*: forming context-specific clusters. For example, depending on the subject-object pair, `"have"` might be grouped along with possessive predicates such as `"has part"`, or with biological relations such as `"grow"` and `"form"`.
(3) *Transfer learning*: having potential to learn new facts, in a similar vein as knowledge base completion models. For instance, if lion can attack zebra, and hyena is clustered with lion, then maybe hyena can also attack zebra.

CoDS COMAD 2020, January 5–7, 2020, Hyderabad, India

ACM ISBN 978-1-4503-7738-6/20/01…$15.00
https://doi.org/10.1145/3371158.3371183

2 SOFT CO-CLUSTERING OF SPO TRIPLES

To capture patterns in the diverse formulation of predicates and subject-object pairs, we compute the clusters jointly and couple them in a meaningful way. For example, given the set of triples: `(student, attack, problem)`; `(researcher, solve, problem)`; `(lion, attack, zebra)`; `(lion, kill, zebra)` and `(lion, kill, antelope)`; we would like to find the following two co-clusters that capture different interpretations of predicates, correctly coupled with the clustered subject-object pairs.

- `{attack, solve}` ↔ `{student-problem, researcher-problem}`
- `{kill, attack}` ↔ `{lion-zebra, lion-antelope}`

For this soft-co-clustering method, we apply non-negative matrix tri-factorization [6] on triples, where matrix rows and columns are associated with subject-object pairs and predicates, respectively. Non-negative matrix factorization is a sophisticated mathematical tool that is extensively used in a large spectrum of applications, such as recommender system, text mining, pattern recognition.

2.1 Non-negative matrix tri-factorization of SPO triples

We form the input matrix by preserving the association between subject-object pairs and predicates in triples. Consider a set of triples with m distinct subject-object pairs and n distinct predicates. From these triples, we form an $m \times n$ binary matrix M, where the element $M_{ij} = 1$ if the subject-object pair SO_i of i^{th} row co-occurs with predicate P_j of j^{th} column. Instead of expressing triples in the form of binary matrix, we can also use the confidence scores of extracted triples as the matrix elements. In matrix tri-factorization, the input matrix M is factorized into three factors as follows:

$$M_{m\times n} = U_{m\times k} \times W_{k\times l} \times V^T_{l\times n}$$

where $U \in \mathbb{R}^{m\times k}$ and $V \in \mathbb{R}^{n\times l}$ are cluster indicator matrices for subject-object pairs and predicates; hyper-parameters k and l represents low-rank dimensionalities for the number of subject-object clusters and predicate clusters; and the middle matrix $W \in \mathbb{R}^{k\times l}$ reflects the alignments between the two type of clusters.

Generally, the degree of diversity in predicate phrases differs from the diversity in subject-object pairs. With the tri-factorization method, we are able to reflect this characteristic by using different numbers of clusters for subject-object pairs and predicates. This makes our tri-factorization approach different from singular value decomposition in principle.

To preserve the soft-co-clustering property in matrix factorization problem, we need to consider the orthonormality constraint on factors U and V; the non-negativity constraint on U, W, V; and the objective function that minimizes the data loss in terms of the Frobenius norm. To this end, the optimization problem is defined as follows:

$$\begin{aligned} Minimize \quad & \|M_{m\times n} - U_{m\times k} \times W_{k\times l} \times V^T_{l\times n}\|_F \\ s.t. \quad & U^T U = I, \quad V^T V = I \\ & U, V, W \geq 0 \end{aligned} \tag{1}$$

We can interpret $U_{i\mu}$ as the likelihood of the membership of i^{th} subject-object pair in μ^{th} subject-object cluster. Similarly, $V_{j\nu}$ represents the likelihood of j^{th} predicate belonging to ν^{th} predicate cluster. The alignment of subject-object clusters to predicate clusters is captured by matrix W, where μ^{th} subject-object cluster is linked to the ν^{th} predicate cluster with likelihood $W_{\mu\nu}$. We can control the soft-clustering property using thresholding. Here, we assign SO_i to the μ^{th} subject-object cluster if $U_{i\mu} \geq \theta_\mu$ and assign P_j to the ν^{th} predicate cluster if $V_{j\nu} \geq \lambda_\nu$, where θ_μ and λ_ν are per-cluster threshold parameters, defined as:

$$\theta_\mu = \delta \cdot \max_i U_{i\mu} \quad and \quad \lambda_\nu = \delta \cdot \max_j V_{j\nu}$$

Here, the parameter $\delta > 0$ controls the thresholding dynamically for each cluster. We tune δ based on the empirical perplexity measure of the cluster assignments of subject-object pairs and predicates after thresholding. We compute the perplexity of two factors U and V by considering that the elements of each cluster follow the multinomial probability distribution. Let us consider that the number of subject-object pairs assigned to μ^{th} subject-object cluster is represented by N_μ, then the perplexity of the factor U is computed as follows:

$$Perplexity(U_{m\times k}) = 2^{-\frac{1}{k}\sum_{\mu\in\{1..k\}} \log_2(\prod_{i\in\{1..m\}, U_{i\mu}\geq\theta_\mu} N_\mu! U_{i\mu})}$$

We compute the perplexity of factor V for predicate clusters similarly. A lower value of perplexity reflects a crisper clustering. We vary the threshold δ within the range $(0, 0.5]$, and found an empirically optimal value of $\delta = 0.1$. We also observe that this proposed thresholding method provides crisper cluster assignments than the fixed-rank thresholding method based on the evaluation of perplexity measure.

Intuitively, each predicate cluster should be coupled with only few subject-object clusters, and vice versa. In order to reflect this characteristic, the factor matrices should be sparse. L_1 regularization of factor matrices are commonly considered in the objective function of matrix factorization to enforce sparsity. However, the formulation of objective function with L_1 norm is non-differentiable, and there is no analytic solution for such tri-factorization model. To solve the optimization problem in Equation 1, we adopt stochastic gradient descent (SGD) proposed by Ding et al. [6], given by the following update rules:

$$U_{ij} \leftarrow U_{ij}\sqrt{\frac{(MVW^T)_{ij}}{(UU^T MVW^T)_{ij}}}$$

$$V_{ij} \leftarrow V_{ij}\sqrt{\frac{(M^T UW)_{ij}}{(VV^T M^T UW)_{ij}}}$$

$$W_{ij} \leftarrow W_{ij}\sqrt{\frac{(U^T MV)_{ij}}{(U^T UWV^T V)_{ij}}}$$

Our SGD-based solver initializes the factor matrices U, V, W with a low density of non-zero values, determined by a hyper-parameter ρ that represents the fraction of non-zero elements in these matrices. Due to the multiplicative property of update rules, initializing factors with a lower value of ρ maintains the sparseness of the factored matrices. Therefore, the overall objective function in our proposed factorization model combines both the data loss in decomposition and the sparseness of W after factorization, defined as follows:

$$Maximize \frac{fraction\ of\ zero\ elements\ (W)}{data\ loss\ by\ Equation\ 1}$$

We tune the hyper-parameters k and l and the sparseness ratio ρ by performing a grid search.

3 EVALUATION

We conduct experiments on a Debian server with Intel Xeon(R) CPU (2 cores@3.20GHz) and 500 GB RAM. We evaluate our proposed approach in terms of quality of predicate clusters and coherence of subject-object and predicate clusters jointly. We use the following two real-world datasets in the evaluation.

• **DBpedia**: consists of 7.6 Million triples from DBPedia[1] where subjects and objects of these triples are linked to DBpedia entities, however the predicates are non-canonicalized infobox properties. The dataset has total 14081 predicates and 1.6 Million subjects.

• **Quasimodo**[2]: contains 2.1 Million triples extracted from different Question-Answering forums and search engine query log [18]. These triples have 132K predicates and 51K subjects.

3.1 Baselines

We compare our proposed soft-co-clustering approach to the clustering based on state-of-the-art KB embedding [5] and rule-mining [7].

Clustering using KB-embedding. We train the TransE embedding model on triples and then use k-means algorithm to cluster the predicates based on their trained embedding vectors. The dimension for the embedding vectors is set to 50 according to the trained model; and the number of k-means clusters is set to the optimal number of clusters found based on our proposed algorithm.

Clustering using rule-mining. We implemented the clustering approach proposed by Galárraga et al. [7]. We use AMIE algorithm to mine equivalence relations between the predicates. Then, predicates are clustered together based on the transitive closure of their equivalence graph. This approach requires subjects and objects of the triples to be linked with a canonicalized KB and only DBpedia dataset satisfies this requirement.

3.2 Experimental setup

We divide triples from both datasets into domain specific verticals and apply clustering on each of them, in order to maintain the efficiency of cluster computation. We slice DBpedia dataset into 26 domains based on DBpedia type hierarchy of entities (e.g., activity, place, person, etc.). The five largest domains are *place*, *person*, *work*, *organization*, and *event*, containing average of 242K subjects and 1.03M triples in each domain. Similarly, we slice Quasimido dataset into 49 basic domains based on the WordNet hierarchy [3], where noun sense of the subject from each triple is mapped to WordNet domain (e.g., animals, plants, earth, etc.). The five largest domains are *earth*, *chemistry*, *animal*, *biology*, and *person*, containing average of 3.9K subjects and 198K triples.

During clustering, we leave out SPO triples where their subject-object pairs co-occur with only one single predicate. We perform grid search of the hyper-parameters: # SO-clusters (k) and # P-clusters (l) are varied within $[50, min(m, n)]$ with $step = 50$ (m and n are the numbers of unique SO-pairs and predicates, respectively); parameter ρ is varied within $[0.1, 0.5]$. Table 1 presents optimal values for hyper-parameters and cluster-specific statistics of our soft-co-clustering method. In particular, we shows the number of triples (#SPO); optimal values for # SO-clusters (k), # P-clusters (l), sparseness ratio (ρ); the average number of elements per SO/P-cluster; and the average number of clusters that a predicate belongs to, indicating multiple interpretations of a predicate. Additionally, we provide macro-averaged statistics for all domains of both datasets. Table 2 shows anecdotal examples of co-clusters for illustration.

Runtime. Co-clustering each slice of Quasimodo and DBpedia take an average of 3 hours and 49 mins, respectively. Clustering predicates including learning time of the predicate embedding using TransE takes an average of 28 mins and 3 hours per slice of Quasimodo and DBpedia, correspondingly. Meanwhile, rule mining approach runs over 6 hours per slice of DBpedia.

3.3 Evaluation of quality of predicate clusters

For Quasimodo, since ground truth is not available, we evaluate the quality of clustering predicates over a set of clusters that are randomly sampled from the output. In contrast, DBpedia provides a partial mapping from infobox raw properties to canonicalized DBpedia Ontology. However, we observe that such mappings are available for only 703 out of total 14K raw predicates and 92% of them are mapped to a unique canonicalized predicate. Therefore, for both datasets, we sample 150 clusters from the clustering output that include at least two predicates. Then, we manually create the ground truth from them. To do this, for each sampled cluster, we consider its most frequent predicate as the representative predicate and the ground truth is created by assembling all other predicates that reflect semantically similar meaning to the representative. For instance, consider the cluster `{birthplace, birthlocation, born at, born during}` with representative predicate `"birthplace"`, then the ground truth is `{birthplace, birthlocation, born at}`.

As we consider only the predicates from the sampled clusters in the construction of ground truth, we focus only on precision measures for evaluation. For example, the precision of the above example cluster is 0.75. Table 3 and Table 4 show the comparison between our proposed soft-co-clustering method and baselines based on three well-established precision measures. We report *macro-precision* that represents the average of the precision of each cluster. Additionally, as a predicate can belong to multiple ground truth clusters, we use two existing co-reference precision measures from literature, B^3*-precision* [2] and ϕ_3*-CEAF-precision* [11]. With all three measures, our proposed soft-co-clustering approach outperforms significantly over the baselines. As DBpedia uses template-based extraction of raw predicates, the diversity and noise in its construction are comparatively less than that of Quasimodo. Reflecting this characteristic, the evaluations show that our method performs better on DBpedia than Quasimodo.

3.4 Evaluation of cluster coherence

To evaluate jointly clustered predicates, subject-object pairs and their alignment, we measure cluster coherence using an intruder

[1]https://wiki.dbpedia.org/downloads-2016-10#datasets
[2]https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/commonsense/quasimodo/

Table 1: Optimal hyper-parameters and statistics for SO-clusters and P-clusters

Dataset	Domain	#SPO	k	l	ρ	# SO/SO-cluster avg.	# SO/SO-cluster max	# P/P-cluster avg.	# P/P-cluster max	# P-clusters/P avg.
Quasimodo	Animal	201942	3500	2000	0.10	38.46	383	2.8	24	1.5
	Person	218924	5000	2000	0.10	11.7	235	4.7	67	1.5
	Sport	30794	1500	400	0.15	13.3	73	3.8	15	1.14
macro-avg. (over	all 49 domains)		1457.8	603.7	0.12	33.97	123.0	3.5	24.8	1.24
DBpedia	Person	1255840	300	100	0.15	45.4	158	2.22	22	1.4
	Place	2007867	700	300	0.15	143.3	641	1.16	9	1.44
	Sport-season	1080750	150	50	0.15	76.1	319	1.82	80	1.13
macro-avg. (over	all 26 domains)		190.6	71.3	0.18	123.31	335.4	1.27	6.93	1.39

Table 2: Anecdotal examples of coupled SO-clusters and P-clusters from both datasets.

Dataset	Clusters of predicates	Clusters of subject-object pairs
DBpedia	before, preceded	Philip I of Castile - Joanna of Castile, Louis I of Naples - Joanna I of Naples, Albert VI Duke of Mecklenburg - Magnus II Duke of Mecklenburg
DBpedia	headquarters, depots, foundingLocation	Orléans Express-Quebec, CK Transit - Ontario, TransBunbury - Bunbury Western Australia
Quasimodo	make noise at, be loud at, make noises at, croak in, croak at, quack at	fox-night, frog-night, rat-night, mouse-night, swan-night, goose-night, chicken-night, sheep-night, donkey-night, duck-night, crow-night
Quasimodo	help in, help with, play part in	butterfly-environment, bee-ecosystem, butterfly-reproduction, butterfly-reproduction of plants, worm-ecosystem

Table 3: Comparison of clustering methods on DBpedia.

Method	*macro-prec.*	B^3*-prec.*	ϕ*-CEAF-prec.*
Soft-co-clustering	0.86	0.98	0.85
KB-embedding	0.52	0.27	0.42
rule-mining	0.77	0.41	0.51

Table 4: Comparison of clustering methods on Quasimodo.

Method	*macro-prec.*	B^3*-prec.*	ϕ*-CEAF-prec.*
Soft-co-clustering	0.80	0.62	0.72
KB-embedding	0.26	0.12	0.27

task. We use a set of 100 randomly sampled aligned SO-clusters and predicate clusters (with at least three elements) from each dataset for the intruder task. In this task, we show annotators sampled subject-object pairs from each cluster along with three predicates from the aligned predicate cluster mixed with a random intruder predicate from the complete data. Then annotators are asked to find the random intruder. For instance, we show SO-cluster `{lion-zebra, hyena-rabbit, cheeta-zebra}` with the predicates `{attack, kill, chase, habitat}`, where `habitat` is the intruder to be found. We design this task in Figure Eight crowdsourcing platform and each task is performed by three annotators. Overall, we get a moderate inter-annotator agreement with Fleiss' kappa value of 0.48. We obtained an intruder detection accuracy of 90% and 65% for DBpedia and Quasimodo, respectively, comparing to 25% for the random baseline. This is a supporting evidence that our co-clustering method yields fairly coherent groups.

4 RELATED WORK

Canonicalizing triples from open information extraction is an integral part of populating KBs. Resolution of named entities is a well-studied problem in the semantic community, often using entity linking methods where extracted mentions of noun phrases are linked to entities in existing KBs [8, 15, 20]. These methods have certain drawbacks for tail entities and emerging entities. To deal with this problem, a well-known approach is to express out-of-KB entities as a set of weighted key phrases [10], and use confidence-based thresholding to recognize emerging entities [10, 15]. However, none of the existing works tackles the issue of canonicalizing emerging entities. Generally, different similarity measures between candidate entities (e.g., string similarity, canopy overlap, etc.) are used in hierarchical clustering to group entities into real-world concepts [7, 19]. The similar approach has been adopted to cluster verb phrases in Resolver system [24]. Galárraga et al. [7] propose a rule-based approach to group equivalent predicates, however, it requires subjects and objects of triples to be linked to an existing KB. With the canonicalized subjects and objects, Nakashole et al. [13] propose PATTY, a subsumption taxonomy of predicates, exploiting the type hierarchy of subjects and objects. This PATTY dictionary is used further to canonicalize predicate phrases [15]. An ILP-based clustering technique has been adopted in TupleKB [12], which enforces hard clustering of predicates. On the other hand, Tandon et al. [22] propose to use a relaxed clustering approach based on word sense disambiguation. Another line of works explores link prediction for KB using KB embedding [5, 9, 16]. Using existing KB embeddings [9], Vashisth et al. [23] canonicalize noun phrases and relations using HAC. As one of the baselines, we also group predicates by applying k-means over the low-dimension vector representation of predicates, learned from KB-embedding [5].

5 CONCLUSION

By adapting the non-negative tri-factorization technique, we have proposed a soft-co-clustering method to jointly cluster subject-object pairs and predicates from a KB and align them in a meaningful manner, for the purpose of canonicalizing triples extracted from open information extraction. The experimental evaluation confirms that our proposed approach creates coherent clusters more efficiently compared to two existing state-of-the-art baselines.

REFERENCES

[1] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007.* 722–735.

[2] Amit Bagga and Breck Baldwin. 1998. Algorithms for Scoring Coreference Chains. In *In The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference.* 563–566.

[3] Luisa Bentivogli, Pamela Forner, Bernardo Magnini, and Emanuele Pianta. 2004. Revising the Wordnet Domains Hierarchy: Semantics, Coverage and Balancing. In *COLING.*

[4] Kurt D. Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008.* 1247–1250.

[5] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.* 2787–2795.

[6] Chris H. Q. Ding, Tao Li, Wei Peng, and Haesun Park. 2006. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, August 20-23, 2006.* 126–135.

[7] Luis Galárraga, Geremy Heitz, Kevin Murphy, and Fabian M. Suchanek. 2014. Canonicalizing Open Knowledge Bases. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014.* 1679–1688.

[8] Ben Hachey, Will Radford, Joel Nothman, Matthew Honnibal, and James R. Curran. 2013. Evaluating Entity Linking with Wikipedia. *Artif. Intell.* 194 (2013), 130–150.

[9] Vinh Thinh Ho, Daria Stepanova, Mohamed H. Gad-Elrab, Evgeny Kharlamov, and Gerhard Weikum. 2018. Rule Learning from Knowledge Graphs Guided by Embedding Models. In *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference, Monterey, CA, USA, October 8-12, 2018, Proceedings, Part I.* 72–90.

[10] Johannes Hoffart, Yasemin Altun, and Gerhard Weikum. 2014. Discovering emerging entities with ambiguous names. In *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014.* 385–396.

[11] Xiaoqiang Luo. 2005. On Coreference Resolution Performance Metrics. In *HLT/EMNLP 2005, Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 6-8 October 2005, Vancouver, British Columbia, Canada.* 25–32.

[12] Bhavana Dalvi Mishra, Niket Tandon, and Peter Clark. 2017. Domain-Targeted, High Precision Knowledge Extraction. *TACL* 5 (2017), 233–246.

[13] Ndapandula Nakashole, Gerhard Weikum, and Fabian M. Suchanek. 2012. PATTY: A Taxonomy of Relational Patterns with Semantic Types. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL 2012, July 12-14, 2012, Jeju Island, Korea.* 1135–1145.

[14] Roberto Navigli. 2009. Word sense disambiguation: A survey. *ACM Comput. Surv.* 41, 2 (2009), 10:1–10:69.

[15] Dat Ba Nguyen, Abdalghani Abujabal, Khanh Tran, Martin Theobald, and Gerhard Weikum. 2017. Query-Driven On-The-Fly Knowledge Base Construction. *PVLDB* 11, 1 (2017), 66–79.

[16] Maximilian Nickel, Lorenzo Rosasco, and Tomaso A. Poggio. 2016. Holographic Embeddings of Knowledge Graphs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.* 1955–1961.

[17] Jay Pujara, Hui Miao, Lise Getoor, and William W. Cohen. 2013. Knowledge Graph Identification. In *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I.* 542–557.

[18] Julien Romero, Simon Razniewski, Koninika Pal, Jeff Z. Pan, Archit Sakhadeo, and Gerhard Weikum. 2019. Commonsense Properties from Query Logs and Question Answering Forums. In *Proceedings of the 28th ACM International Conference on Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019.*

[19] Alieh Saeedi, Eric Peukert, and Erhard Rahm. 2018. Using Link Features for Entity Clustering in Knowledge Graphs. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings.* 576–592.

[20] Wei Shen, Jianyong Wang, and Jiawei Han. 2015. Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions. *IEEE Trans. Knowl. Data Eng.* 27, 2 (2015), 443–460.

[21] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007.* 697–706.

[22] Niket Tandon, Gerard de Melo, Fabian M. Suchanek, and Gerhard Weikum. 2014. WebChild: harvesting and organizing commonsense knowledge from the web. In *Seventh ACM International Conference on Web Search and Data Mining, WSDM 2014, New York, NY, USA, February 24-28, 2014.* 523–532.

[23] Shikhar Vashishth, Prince Jain, and Partha Talukdar. 2018. CESI: Canonicalizing Open Knowledge Bases using Embeddings and Side Information. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018.* 1317–1327.

[24] Alexander Yates and Oren Etzioni. 2009. Unsupervised Methods for Determining Object and Relation Synonyms on the Web. *J. Artif. Intell. Res.* 34 (2009), 255–296.