When Lexicon-Grammar Meets Open Information Extraction: a Computational Experiment for Italian Sentences

Raffaele Guarasci, Emanuele Damiano, Aniello Minutolo, Massimo Esposito
National Research Council of Italy
Institute for High Performance Computing and Networking (ICAR), Naples, Italy
{name.surname}@icar.cnr.it

Abstract

In this work we show an experiment on building an Open Information Extraction system (OIE) for Italian language. We propose a system wholly reliant on linguistic structures and on a small set of verbal behavior patterns defined putting together theoretical linguistic knowledge and corpus-based statistical information[1]. Starting from elementary one-verb sentences, the system identifies elementary tuples and then, all their permutations, preserving the overall well-formedness (grammaticality) and trying to preserve semantic coherence (acceptability). Although the work focuses only on the Italian language, it can be proficiently extended also to other languages, since it is essentially based only on linguistic resources and on a representative corpus for the language under consideration[2].

1 Introduction

One of the most interesting approach to handle the rapid growth of textual data emerged in the last decade is Open Information Extraction (OIE). Starting from natural language sentences, it allows to extract one or more domain-independent propositions, scaling to the diversity and size of the corpus considered (Banko et al., 2007). Each extracted proposition is represented by a verb and its arguments, i.e. "Maria goes to the party" is a proposition with a relation (the verb *goes*) that links together two arguments (*Maria*, *the party*). Arguments (nouns or noun groups) can have different roles (subject, direct object…) and they can be mandatory or optional. In this sentence, both arguments *Maria* (subject) and *the party* (direct object) are mandatory, so it is impossible to remove one of them or the sentence becomes unacceptable from a grammatical point of view. Due to the high field of Natural Language Processing (NLP) tasks in which OIE outputs can be used (Christensen et al., 2013; Fader et al., 2014; Stanovsky et al., 2015; 2016; Khot et al., 2017; Rahat et al., 2017), numerous OIE approaches for English have been developed. However, being a language-dependent task, OIE systems cannot be shifted from one language to another, i.e. a system created for English is not compatible with Italian. Moreover, many of the proposed OIE approaches rest on unstable grounds. Some of them use heuristics to manage large quantities of textual data, others lack the support of a theoretical basis, outlining the natural language in a reductive way.

Differently from the vast majority of existing OIE approaches, we propose a linguistic-based unsupervised system designed to extract n-ary propositions (not only "relation-argument" triples) from natural language sentences in Italian, ensuring domain independence and scalability.

Our system aims to identify the elementary tuple(s) from the input sentence, then all its (their) permutations, by adding progressively arguments composing the sentence. After that – according the behavior patterns of the verb – it generates every possible syntactically valid n-ary proposition, granting grammaticality.

To reach this result we have combined two types of resources. To gather information about verb behavior in sentences, we grounded our work on the linguistic basis provided by Lexicon Grammar (LG) (Gross, 1994). In order to obtain a fine-grained characterization of arguments, we

[1] An online demo showing some features of the system is freely available at the address https://nlpit.na.icar.cnr.it/

combine this theoretical knowledge with distributional corpus-based information extracted from itWaC (Baroni et al., 2009). From LG tables we extract patterns of verbs behaviors, and from itWaC we enrich these patterns with statistical information. Using complex linguistic structures and dependency parse trees (DPT) we can detect verbal behavior patterns occurring in one-verb sentences and generate from them all the possible well-formed propositions, by adding complements and adverbials. The use of formal patterns derived from a theoretical framework allows to better distinguish between necessary verbal arguments and optional removable adjuncts and to verify syntactic restrictions in verb possible structures.

Arguments optionality and syntactic constraints are critical features to grant the grammaticality of the propositions generated, also trying to approximate a first level of semantic acceptability.

2 Related Work

In the last years, several approaches to OIE has been developed (Banko et al., 2007; Zhu et al., 2009; Wu et al., 2010; Fader et al., 2011; Schmitz et al., 2012; Del Corro et al., 2013), all of them with the characteristic of utilizing a set of patterns in order to obtain propositions, granting scalability and portability across different domains.

They differ in many aspects such as performances (precision, recall, speed); linguistic structures used (Part-of-Speech tags, chunks, DPT); patterns to extract information (hand-crafted based on heuristics or learned from a training corpus); type of generated output (binary extractions, n-ary extractions, nested extractions).

However, most of these existing approaches so far has been focused on English, with only some recent attempts that have appeared for other languages, such as Spanish (Zhila et al, 2013), Chinese (Wang et al, 2014), Vietnamese (Truong et al., 2017), German (Falke et al., 2016; Bassa et al., 2018) and Romance languages (Gamallo et al., 2012; Gamallo et al., 2015). As far as we know only one approach has been attempted for the Italian (Damiano et al., 2018). It is a preliminary experiment based on a limited set of patterns and heuristics, and experimented on a hand-crafted dataset of reduced size.

3 Lexicon-Grammar

As the theoretical basis for our system we decided to use LG since it regards the systematic formalization of a very broad quantity of data for the Italian language (Elia et al., 1981; D'Agostino, 1992). Other resources describing a subset of Italian verbs have been developed, such as LexIt (Lenci et al. 2012), MultiWordNet (Pianta et al. 2002), SensoComune (Oltramari et al. 2013) and T-PAS (Jezek et al., 2014). However, none of them provides a formal classification of verbs in classes or clusters. Conversely, LG groups verbs in classes according to their behavior, specifying for each verb its essential arguments and possible syntactic structures in order to create well-formed sentences (Leclère, 2002).

3.1 How data are structured in LG

LG classes are represented in the form of tables. Each row of the table corresponds to a verb of the class, each column lists all properties that may be valid or not for the different members of the class. At the intersection of a row and a column, the symbol **+** or **–** may indicate that the property corresponding to the column is valid or not for the verb corresponding to the row, as shown in Table 1[3], which reports some Italian verbs and their properties as encoded in a LG. Properties can be of different types. They can refer to the syntactic structure and the prepositions admitted by that specific verb, semantic restrictions (e.g. human/non-human argument) or possible transformations (e.g. passive form). For the purpose of this work, only syntactic properties will be considered. This choice reflects the syntactic nature of OIE, which focuses on shapes and structures of verbs.

Verb	N_0VN_1	N_0V	$N_0VprepN_1$	$N_0VN_1prepN_2$
Mangiare (*to eat*)	+	+	–	–
Muovere (*to move*)	+	-	–	+
Girare (*to turn*)	+	+	+	+

Table 1 Example of an LG table

The first column contains the defining property, which corresponds to the basic syntactic structure

[3] The formal notation used in LG is summarized as follows: N indicates a nominal group and is followed by a progressive subscript indicating its nature (N0 is the subject, N1 is the first complement, N2 is the second complement, etc.), V represents the verb, prep indicates prepositions.

of the elementary sentence. The property expressed in the second column is a syntactic property called deletion (Harris, 1982), labeled as N_0V, which allows the cancellation of the element N_1 from the basic syntactic structure specified with the defining property. Deleting the element N_1 on the right of the verb is valid for the verb "mangiare" ("Max mangia", *Max eats*), while it produces ungrammatical unacceptable sentences for the verb "muovere" ("*Max muove", **Max moves*). `Prep` represents a set of every possible adjuncts placed before every argument N_i.

3.2 From tables to patterns

Despite the richness of this fine-grained information, LG tables suffer from some limitations that have made them useless in real NLP applications: they are verbose and properties is neither uniform nor standardized. Therefore, many changes were necessary to be able to use these resources in the OIE system:

Grouping. We divided verbs into classes: direct (`D`) without preposition, indirect with a preposition (`I`), and locative (`L`). This distinction is preferred to the classical distinction between transitive and intransitive verbs, since locative verbs can accept both transitive and intransitive construction. Verbs assuming a copulative function (support verbs) form a further class (`S`). For the purpose of this work, we do not consider complement-clause verbs, because of the variability of the structures possible for the definition of unique patterns.

Enrichment: `Prep` element is too coarse. We need to specify which kind of preposition the selected verb admits. To overcome this limit, we add a syntactic profile to each verb, containing the most frequent prepositions associated to it. We extract this information from itWaC corpus.

Formal representation. To reduce redundant information of the original tables we formalize a grammar to compactly represent verbs behavior, indicating selection preferences on the possible arguments of a verb. Square brackets `[ ]` represent the possibility of deleting arguments, round brackets `( )` indicates there are many possible arguments separated by a vertical bar, and `XOR` symbol ⊕ represents the exclusive alternativity of patterns.

As it is shown table 2, the notation $N_0V[N_1]$ indicates that the verb "mangiare" (*to eat*) can accept both the structures N_0VN_1 or N_0V, and the notation $N_0V(in|a)N_1$ denotes that the verb can accept alternatively and also simultaneously both the patterns N_0VinN_1 and N_0VaN_1. On the other hand, a notation like $N_0VN_1 \oplus N_0VinN_1$ denotes that the verb can accept exclusively only one between the patterns N_0VN_1 and N_0VinN_1, even if they are both valid from a grammatical perspective. This is due to the fact that their selection preferences are representative of different verb usages and, thus, are alternative and exclusive from a semantic perspective. Note that in the table 2 possible prepositions are reduced for a better readability of the pattern.

Verbs	Patterns
mangiare	$N_0V[N_1]$
(*to eat*)	
muovere	$N_0VN_1 \oplus$
(*to move*)	$N_0V(in<\textit{in}> \mid da<\textit{from}> \mid verso<\textit{toward}>)N_1$
girare	$N_0V(a<\textit{to}> \mid intorno<\textit{around}>)N_1 \oplus$
(*to turn*)	$N_0VN_1[(a<\textit{to}> \mid da<\textit{from}> \mid verso<\textit{toward}>)N_2]$

Table 2 Patterns derived from LG tables

4 Proposed Approach

Our approach for OIE is arranged in the form of a multi-step pipeline and it consists into 4 steps:

Sentence Processing*:* every input sentence is checked to verify that it is suitable for the approach.

Arguments Identification*:* arguments of the verb are identified (i.e. subjects, direct complements, indirect complements…).

Pattern Recognition: verbal structures that match the patterns are identified and elementary tuples made by the combination of arguments are generated.

Proposition Generation: n-ary propositions depending on the elementary tuples and the remaining arguments (i.e. adverbs, complements and modifiers) are generated.

As an example, for the sentence "Da domani Anna andrà da Roma a Milano" (*From tomorrow Anna will go from Rome to Milan*), both the tuples and corresponding propositions that are generated are reported in Table 3.

The verb "andare" (*to go*) belongs to locative group loc, and its complete pattern is the following $N_0V[daN_1](a|in|verso|su|sopra)N_2$. In the first column of the table identified patterns for the verb are reported, the second column lists tuples and propositions generated from every single pattern.

Pattern	Generations
N_0VaN_1	1. (“Anna”<*Anna*>, “andrà”<*will go*>, “Milano”<*Milan*>) Anna andare a Milano (*Anna to go to Milan*) 2. (“Domani”<*tomorrow*>, “Anna”<*Anna*>, “andrà”<*will go*>, “Milano”<*Milan*>) Da domani Anna andare a Milano (*From Tomorrow Anna to go to Milan*)
N_0daVaN_1	3. (“Anna”<*Anna*>, “andrà”<*will go*>,”Roma”<*Rome*>,“Milano”<*Milan*>) 4. Anna andare da Roma a Milano (*Anna to go from Rome to Milan*) (“Domani”<tomorrow>, “Anna”<Anna>, “andrà”<will go>, “Roma”<*Rome*>,“ “Milano”<Milan>) Da domani Anna andare da Roma a Milano (*From Tomorrow Anna to go from Rome to Milan*)

Table 3 tuples and propositions generated from an input sentence

5 Experiment and validation

We carried out the evaluation using quantitative metrics well known in NLP literature: precision and recall. Precision measures the average on all the sentences of the percentage of extractions obtained by the proposed approach that are correct, whereas recall measures the average on all the sentences of the percentage of extractions manually annotated in the dataset that are correctly identified by the proposed approach. Performances was evaluated on a dataset of sentences containing verbs belonging to different classes, and the validation took place with respect to grammaticality and acceptability (i.e. syntactic well-formedness of the sentences and its meaningfulness in the context) using the gold standard proposed in (Guarasci et al. *in press*). Notice that grammaticality and acceptability judgements is a much debated topic in theoretical and computational linguistics in the past (Phillips, 2009; Phillips, 2011; Gibson et al., 2010) and still today it is considered a controversial subject (Lau et al., 2017; Sprouse et al.; 2018). Even if OIE is a syntactic task, so it focus on the structure of the sentence, but not its meaning (Lau et al., 2017), we aim to generate sentences not only well-formed but also respecting some syntactic constraints and selection preferences, trying to approximate the first level of semantic acceptability.

	Sentences	Grammaticality		Acceptability	
		P	**R**	**P**	**R**
Total verbs	195	0.91	0.78	0.79	0.84
Locative	62	0.93	0.73	0.77	0.83
Direct	30	0.90	0.93	0.79	0.93
Indirect	65	0.88	0.81	0.78	0.83
Support	38	0.98	0.66	0.86	0.78

Table 4 results for different verb classes

Table 4 shows precision (P) and recall (R) scores with respect to the two criteria on the verbs divide by classes.
Precision and recall achieve high values with respect to both grammaticality and acceptability. More precisely, with respect to the different structures of verbs considered, precision has resulted sensibly higher for sentences containing support verbs with respect to grammaticality and acceptability. This behavior is reversed for recall, which has resulted for sentences containing direct, indirect or locative verbs.

5.1 Comparison with other OIE systems

Globally, generations per sentences and performances achieved are comparable with state-of-the-art OIE systems in other languages, respectively ClausIE (English) and GerIE (German). Moreover, we compare our results with the only other experiment conducted on Italian presented by the authors and named ItalIE (Damiano et al, 2018).

	Sentences	Grammaticality		Acceptability	
		P	**R**	**P**	**R**
Total verbs	195	0.84	0.40	0.73	0.43
Locative	62	0.91	0.46	0.74	0.51
Direct	30	0.82	0.56	0.74	0.57
Indirect	65	0.72	0.27	0.68	0.57
Support	38	0.91	0.36	0.86	0.45

Table 5 Performances of ItalIE

As shown in Tables 5, our approach has reached the best overall performances in terms precision and recall for both grammaticality and acceptability. ItalIE highlighted a sensibly lower number of generations (511 vs 918 of our approach) with a moderate decrease in precision but a significant reduction in recall. This behavior can be explained by the fact that ItalIE is based on a fixed set of clause patterns not considering the extreme variability of verb behaviors and also the selection preferences on their possible arguments. Furthermore, its algorithm based on DPT to identify constituents through dependency relations has shown some weaknesses. It fails in detecting and properly handling named entities, multi-word expressions, adjectives, numerals, dates and some patterns related to support verbs.

5.2 Error Analysis

The number of both false positives and negatives generated in the experiments is shown in Table 6 with respect to grammaticality (G) and acceptability (A).

	False positives					False negatives		
	DP	**NE**	**SC**	**MC**	**Tot**	**DP**	**VU**	**Tot**
G	78	3	0	0	81	145	86	231
A	78	3	76	38	195	114	21	135

Table 6 Summary of the errors generating false positives and negatives with respect to grammaticality and acceptability.

Various types of errors are divided as follows:

DP: errors caused by incorrect dependency parsing due to wrong and/or missing dependencies between element occurring in the input sentence. They represent the vast majority of the errors affecting overall performances of the proposed approach. With respect to grammaticality and acceptability, false positives have been generated by DP errors in 96% and 40% of cases, whereas false negatives are due to DP errors in 63% and 84% of cases, respectively.

NE: error in the identification of named-entities. NE errors have occurred in a not significant number of cases, only 3, generating false positives with respect to both grammaticality and acceptability.

VU: behavior patterns not associated to the verb usage selected for the input sentence. It represents the second source of errors causing false negatives with respect to grammaticality and acceptability (in 37% and 16% of cases, respectively).

MC: missing morpho-syntactic concordance among different parts-of-speech or missing contractions or combinations between prepositions and articles. It causes 19% of false positives in acceptability.

SC: violated semantic constraints. It affects only acceptability, causing 39% of false positives. Notice that this error is referred only to the semantic perspective, while others are related to grammatical aspects.

6 Conclusions and Future Work

In this work we have shown an experiment to perform OIE for Italian language, extracting n-ary propositions from natural language sentences, granting well-formedness of the generations. The system relies on a linguistic resource (LG) and on a representative corpus for Italian (itWaC). While these resources are specific to Italian, they also exist for other languages, so the system can be easily extended. In particular, LG tables exist in digital format also for French (Tolone, 2012), English (Garcia-Vega, 2010; Machonis, 2010), Portuguese (Baptista, 2001), Romanian (Ciocanea, 2011). Likewise, the itWaC corpus used in this work is part of the WaCky Wide Web corpora collection (Baroni et al., 2009), which includes corpora of English (ukWaC), German (deWaC), French (frWac). Concerning performances of the system, although the results are encouraging, we are looking forward to further developments.

With regard to methodological progress, we plan to integrate novel methods based on deep learning to increase the performance of the system, trying to reduce DP errors and better handle named entities, frozen and semi-frozen bigrams and multiword expressions. From an applicative perspective, this work will be experimented in Italian Question Answering system, with the goal to improve the ability in reading complex texts and extracting the correct answers to users' questions. Other possible outcomes can include text summarization or other NLP tasks.

References

Michele Banko, Michael J Cafarella, Stephen Soderland, Matthew Broadhead, and Oren Etzioni. 2007. Open information extraction from the web. In *Proceeding of IIJCAI*, vol. 7, pp. 2670-2676.

Jorge Baptista. 2012. Viper: A lexicon-grammar of european portuguese verbs. In *31e Colloque International sur le Lexique et la Grammaire*.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources & Evaluation*, 43(3):209–226, September.

Akim Bassa, Mark Kröll, and Roman Kern. 2018. GerIE-An Open Information Extraction System for the German Language. *Journal of Universal Computer Science*, 24(1):2–24.

Janara Christensen, Stephen Soderland, and Oren Etzioni. 2013. Towards Coherent Multi-Document Summarization. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp 1163–1173

Cristiana Ciocanea. 2011. *Lexique-grammaire des constructions converses en a da/ a primi en roumain. (Lexicon-grammar of converse constructions in a da/ a primi in Romanian)*. PhD Thesis, University of Paris-Est, France.

Emilio D'Agostino. 1992. *Analisi del discorso: metodi descrittivi dell'italiano d'uso*. Loffredo.

Emanuele Damiano, Aniello Minutolo, and Massimo Esposito. 2018. Open Information Extraction for

Italian Sentences. In *Proceedings of 2018 32nd International Conference on Advanced Information Networking and Applications Workshops*, pp. 668-673

Luciano Del Corro and Rainer Gemulla. 2013. ClausIE: clause-based open information extraction. In *Proceedings of the 22nd International Conference on World Wide Web*, pp. 355–366.

Annibale Elia, Maurizio Martinelli, and Emilio d'Agostino. 1981. *Lessico e strutture sintattiche: introduzione alla sintassi del verbo italiano*. Liguori, Napoli.

Oren Etzioni, Anthony Fader, Janara Christensen and Stephen Soderland. 2011. Open Information Extraction: The Second Generation. In *Proceeding of IJCAI*, vol. 11, pp. 3-10.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying Relations for Open Information Extraction. In *EMNLP '11*, pages 1535–1545, Stroudsburg, PA, USA. Association for Computational Linguistics.

Anthony Fader, Luke Zettlemoyer, and Oren Etzioni. 2014. Open question answering over curated and extracted knowledge bases. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1156-1165.

Pablo Gamallo, Marcos Garcia, and Santiago Fernández-Lanza. 2012. Dependency-based Open Information Extraction. In *ROBUS-UNSUP '12*, pages 10–18, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pablo Gamallo, Marcos Garcia. 2015. Multilingual open information extraction. In *Portuguese Conference on Artificial Intelligence*, pp. 711-722.

Edward Gibson and Evelina Fedorenko. 2013. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1–2):88–124.

Maurice Gross. 1994. *Constructing lexicon-grammars*. Centre national de la recherche scientifique, Universités de Paris 7 et 8.

Zellig Sabbettai Harris. 1982. *A grammar of English on mathematical principles*. John Wiley & Sons Incorporated.

Elisabetta Jezek, Bernardo Magnini, Anna Feltracco, Alessia Bianchini, and Octavian Popescu. T-PAS: A resource of corpus-derived Typed Predicate Argument Structures for linguistic analysis and semantic processing. In *Proceedings of LREC*, pp. 890-895.

Tushar Khot, Ashish Sabharwal, and Peter Clark. 2017. Answering complex questions using open information extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, (2)pp. 311-316.

Jey Han Lau, Alexander Clark, and Shalom Lappin. 2017. Grammaticality, Acceptability, and Probability: A Probabilistic View of Linguistic Knowledge. *Cognitive Science*, (41): 5, pp. 1202-1241.

Christian Leclère. 2005. The Lexicon-Grammar of French Verbs. In *Linguistic Informatics State of the Art and the Future: The first international conference on Linguistic Informatics*, (1)pp. 29-45.

Christian Leclère. 2002. Organization of the lexicon-grammar of French verbs. *Lingvisticæ Investigationes*, 25(1):29–48, January.

Alessandro Lenci, Gabriella Lapesa, and Giulia Bonansinga. LexIt: A Computational Resource on Italian Argument Structure. In *LREC*, pp. 3712-3718.

Alessandro Oltramari, Guido Vetere, Maurizio Lenzerini, Aldo Gangemi, and Nicola Guarino. 2010. Senso Comune. In *LREC* pp. 3873-3877.

Colin Phillips. 2009. Should we impeach armchair linguists. *Japanese/Korean Linguistics*, 17:49–64.

Colin Phillips. 2013. Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes*, 28(1–2):156–187.

Emanuele Pianta, Luisa Bentivogli, and Christian Girardi 2002. Developing an aligned multilingual database. In *Proceedings of Global WordNet Conference*.

Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni. 2012. Open Language Learning for Information Extraction. In *EMNLP-CoNLL '12*, pages 523–534, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jon Sprouse, Beracah Yankama, Sagar Indurkhya, Sandiway Fong, and Robert C Berwick. 2018. Colorless green ideas do sleep furiously: gradient acceptability and the nature of the grammar. *The Linguistic Review*, 35(3):575–599.

Gabriel Stanovsky and Ido Dagan. 2015. Open IE as an Intermediate Structure for Semantic Tasks. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2300-2305.

Gabriel Stanovsky and Ido Dagan. 2016. Creating a large benchmark for open information extraction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2300–2305.

Elsa Tolone. 2012. Analyse syntaxique à l'aide des tables du Lexique-Grammaire du français. *Lingvisticæ Investigationes*, 35(1):147–151.

Diem Truong, Duc-Then Vo, Uyen Trang Nguyen. 2017. Vietnamese Open Information Extraction. In *Proceedings of the Eighth International Symposium on Information and Communication Technology*, pp. 135-142.

Mingyin Wang, Lei Li, and Fang Huang. 2014. Semi-supervised chinese open entity relation extraction. In *2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems*, pages 415–420.

Fei Wu and Daniel S Weld. 2010. Open Information Extraction using Wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 118–127.

Alisa Zhila and Alexander Gelbukh. 2013. Comparison of open information extraction for English and Spanish. *Computational Linguistics and Intelligent Technologies*, 12(19):714–722.

Jun Zhu, Zaiqing Nie, Xiaojiang Liu, Bo Zhang, and Ji-Rong Wen. 2009. StatSnowball: a statistical approach to extracting entity relationships. In *Proceedings of the 18th international conference on World wide web*, pages 101–110. ACM.