

Utilisation de Représentations Distribuées de Relations pour la Désambiguïsation d'Entités Nommées

Nicolas Wagner Romaric Besançon Olivier Ferret

CEA, LIST, Laboratoire Vision et Ingénierie des Contenus, Gif-sur-Yvette, F-91191, France.

{romaric.besancon,olivier.ferret}@cea.fr

RÉSUMÉ

L'identification des entités nommées dans un texte est une étape fondamentale pour de nombreuses tâches d'extraction d'information. Pour avoir une identification complète, une étape de désambiguïsation des entités similaires doit être réalisée. Celle-ci s'appuie souvent sur la seule description textuelle des entités. Or, les bases de connaissances contiennent des informations plus riches, sous la forme de relations entre les entités : cette information peut également être exploitée pour améliorer la désambiguïsation des entités. Nous proposons dans cet article une approche d'apprentissage de représentations distribuées de ces relations et leur utilisation pour la tâche de désambiguïsation d'entités nommées. Nous montrons le gain de cette méthode sur un corpus d'évaluation standard, en anglais, issu de la tâche de désambiguïsation d'entités de la campagne TAC-KBP.

ABSTRACT

Exploiting Relation Embeddings to Improve Entity Linking

The identification of named entities in documents is the basis of most Information Extraction tasks. The full identification of a named entity includes its disambiguation, which generally relies on its textual description only. However, the relations between entities in knowledge bases are also an interesting source of information that can be exploited for performing such disambiguation. In this article, we propose to build word embeddings for representing such relations and to use these representations for disambiguating entity mentions in texts. Furthermore, we show the interest of this approach on the 2015 TAC-KBP corpus, a standard corpus for evaluating Entity Linking systems.

MOTS-CLÉS : Désambiguïsation d'entités nommées, apprentissage de représentations, relations.

KEYWORDS: Entity linking, named entities, embeddings, relations.

1 Introduction

Dans le domaine de l'extraction d'information à partir de textes, la reconnaissance d'entités nommées prend une place particulièrement importante, d'une part parce qu'elle peut servir à représenter le contenu thématique de certains documents de façon informative et concise (de qui on parle, où se situe l'action ...) et d'autre part parce que ces entités sont en général nécessaires à tous les traitements plus complexes qui peuvent être entrepris à la suite, comme l'extraction de relations ou d'événements, ou le remplissage automatique de bases de connaissances.

De ce fait, la qualité et la précision de la reconnaissance des entités nommées est fondamentale. Une étape additionnelle augmentant la précision de l'identification des entités consiste à lever l'ambiguïté

entre des entités ayant des mentions identiques ou similaires. Cette étape de désambiguïsation d'entités nommées (aussi appelée liaison référentielle d'entités ou *Entity Linking*) s'appuie sur l'utilisation d'une base de connaissances contenant les entités connues *a priori* servant de référence pour la désambiguïsation (Shen *et al.*, 2015; Ling *et al.*, 2015). Par exemple, la mention d'entité « *Castro* » correspond à deux entités différentes dans les contextes suivants¹ :

Obama and Castro shake hands as U.S. and Cuba seek better ties.

→ Raúl Castro

Clinton/Castro would blow up Jeb's dream of aligning with the Hispanic voters.

→ Julián Castro

Traditionnellement, cette désambiguïsation s'appuie sur la similarité entre le contexte textuel de la mention de l'entité à désambiguïser et une description textuelle associée aux entités dans la base de connaissances (le plus souvent, le texte de la page Wikipédia associée à l'entité). Or, les bases de connaissances possèdent également une structure relationnelle associant les entités entre elles. Plusieurs études ont montré que ces relations peuvent être exploitées pour aider la désambiguïsation, que ce soit dans un cadre général de désambiguïsation lexicale (Moro *et al.*, 2014) ou dans le cadre plus spécifique de la désambiguïsation des entités nommées (Usbeck *et al.*, 2014), en particulier pour celles dont le contenu textuel associé est pauvre ou inexistant (Besançon *et al.*, 2016).

Par ailleurs, des travaux récents ont montré que l'usage de représentations sémantiques distribuées (*word embeddings*) sur le modèle de *word2vec* (Mikolov *et al.*, 2013), construites pour représenter de façon jointe les mots des textes et les entités nommées d'une base de connaissances, peut améliorer la détection de la similarité entre les mentions d'entités et les entités de cette base, augmentant ainsi la précision de la désambiguïsation (Fang *et al.*, 2016; Yamada *et al.*, 2016; Moreno *et al.*, 2017).

Nous proposons, dans cette étude, d'associer ces deux idées en apprenant des représentations distribuées des relations entre les entités d'une base de connaissances et en exploitant ces représentations pour aider la désambiguïsation des entités nommées.

2 Approche

2.1 Apprentissage de représentations des relations entre entités

La désambiguïsation des entités nommées se fait en utilisant une base de connaissances contenant des entités auxquelles sont associées un ensemble d'informations. Parmi les bases de connaissances traditionnellement utilisées pour cette tâche, on peut citer YAGO (Suchanek *et al.*, 2007), DBPedia (Lehmann *et al.*, 2015), Freebase (Bollacker *et al.*, 2008) ou Wikidata (Vrandečić & Krötzsch, 2014). Les informations associées aux entités prennent la forme de propriétés ou de relations typées entre les entités : par exemple, à une entité de type personne pourra se voir associer son *âge* comme propriété et son *lieu de naissance* comme relation avec une autre entité, de type Lieu, également présente dans la base. Certaines relations peuvent également impliquer plus de deux entités (par exemple, une relation de type *mariage* associe deux personnes, un lieu et une date).

Notre objectif est d'exploiter toutes ces relations entre entités (binaires ou n-aires) existant dans la base de connaissances de référence pour aider la désambiguïsation des entités. Ces relations fournissent en effet une information importante concernant les entités. Pour reprendre l'exemple précédent sur la

1. Les exemples de textes, d'entités et de relations présentés dans l'article sont extraits des données de la campagne d'évaluation TAC-KBP pour la tâche de désambiguïsation d'entités nommées (EDL).

mention « *Castro* », les relations pour les deux entités « *Raúl Castro* » et « *Julián Castro* », présentées dans le tableau 1, montrent des éléments pouvant aider la désambiguïsation : par exemple, le lien avec Cuba pour le premier et celui avec le parti démocrate américain pour le second.

Relation	Raúl Castro	Julián Castro
<i>party</i>	Communist Party of Cuba	Democratic Party
<i>position</i>	Prime Minister of Cuba	Mayor of San Antonio
<i>jurisdiction</i>	Cuba	San Antonio
<i>spouse</i>	vilma espín	Erica Lira Castro
<i>education</i>	Belen Jesuit Preparatory School	harvard law school
<i>sibling</i>	Fidel Castro	Joaquín Castro

TABLE 1 – Exemple de relations pour les entités « *Raúl Castro* » et « *Julián Castro* »

Dans ce but, nous proposons de définir une similarité relationnelle entre une mention d’entité et les entités de la base de connaissances en nous appuyant sur l’apprentissage de représentations jointes entre les mots et les entités, fondées sur leur contexte relationnel, afin de déterminer si une mention et une entité candidate sont compatibles du point de vue de leurs relations.

Plusieurs modèles pour l’apprentissage de représentations de relations (*embeddings* de relations) ont été proposés (Nickel *et al.*, 2012; Riedel *et al.*, 2013; Weston *et al.*, 2013), mais principalement dans le but d’améliorer l’extraction de relations : leur objectif est donc d’estimer la probabilité de l’existence d’une relation entre deux entités dans une base de connaissances. Nous nous plaçons pour notre part dans une optique différente, où nous souhaitons comparer les contextes relationnels de deux éléments. De plus, nous voulons avoir la possibilité d’apprendre, de façon jointe, des représentations pour les mots et les entités en fonction de leur relations. Nous proposons donc d’utiliser un apprentissage de représentations s’appuyant sur *word2vec*, mais en utilisant des contextes relationnels au lieu de contextes textuels, de manière similaire au modèle DeepWalk (Perozzi *et al.*, 2014).

Dans (Moreno *et al.*, 2017), un apprentissage joint de représentations de mots et d’entités est déjà réalisé en exploitant les contextes textuels des entités. Nous proposons ici d’étendre cette idée aux contextes relationnels : à partir du graphe des relations impliquant une entité, ces contextes sont construits comme des séquences d’entités en relation obtenues selon un modèle de marche aléatoire dans ce graphe comparable à celui de DeepWalk. Dans notre cas néanmoins, nous initialisons deux marches aléatoires à partir d’une entité pour construire un contexte gauche et un contexte droit de l’entité, ce qui permet d’avoir un nombre fixe de contextes pour chaque entité.

Plus précisément, soit e une entité de la base de connaissances, associée à un identifiant unique $id(e)$ et un nom $nom(e)$, non nécessairement unique puisque correspondant à la chaîne de caractères d’une mention de e . Soit également $Rel(e) = \{e_1, e_2, e_3, \dots, e_n\}$, l’ensemble des entités en relation avec e . On construit un ensemble de séquences ayant e comme élément central en sélectionnant aléatoirement deux entités e_i et $e_j \in Rel(e)$ et en les plaçant comme contexte immédiat gauche et droit de e , puis en itérant à partir de e_i et e_j . Au deuxième niveau, on obtient donc une séquence du type :

$$e_k \in Rel(e_i) \leftrightarrow e_i \in Rel(e) \leftrightarrow e \leftrightarrow e_j \in Rel(e) \leftrightarrow e_l \in Rel(e_j)$$

À partir de cette séquence, deux contextes sont construits, un utilisant l’identifiant de l’entité, l’autre utilisant le nom de l’entité :

$$\begin{array}{c} \text{nom}(e_k) \text{ nom}(e_i) \mathbf{nom}(e) \text{ nom}(e_j) \text{ nom}(e_l) \\ \text{nom}(e_k) \text{ nom}(e_i) \mathbf{id}(e) \text{ nom}(e_j) \text{ nom}(e_l) \end{array}$$

En reprenant l'exemple du tableau 1 et pour une taille de contexte de 1, on générerait pour l'entité « *Raúl Castro* » (associée à l'identifiant *m.01gcjq*) donc des contextes du type :

<i>Communist Party of Cuba</i>	<i>Raúl Castro</i>	<i>Cuba</i>
<i>Fidel Castro</i>	<i>Raúl Castro</i>	<i>Prime Minister of Cuba</i>
<i>Communist Party of Cuba</i>	<i>m.01gcjq</i>	<i>Cuba</i>
<i>Fidel Castro</i>	<i>m.01gcjq</i>	<i>Prime Minister of Cuba</i>
...		

Un modèle *word2vec* est alors entraîné à partir de ces contextes. Le fait de générer deux pseudo-phrases pour chacune des entités, une avec son nom et l'autre avec son identifiant, permet de créer une représentation jointe d'entités et de mots.

2.2 Similarité entre mention et entité

L'apprentissage de représentation pour les mentions et les entités décrit à la section précédente nous donne la possibilité de comparer une mention et une entité directement par une mesure de similarité entre leurs représentations. Néanmoins, cette comparaison peut être enrichie en prenant en compte le contexte relationnel de la mention d'entité. Ce dernier, appelé dans ce qui suit contexte relationnel textuel, est défini par l'ensemble des mentions d'entités proches de la mention cible dans le texte. Pour la phrase ci-dessus « *Obama and **Castro** shake hands as U.S. and Cuba seek better ties.* » par exemple, le contexte relationnel textuel de *Castro* est ainsi formé par les mentions {*Obama, U.S., Cuba*}. Cette notion constitue le pendant textuel du contexte relationnel d'une entité au sein de la base de connaissances. À la différence de celui-ci cependant, le contexte relationnel textuel n'est pas fondé sur des relations typées explicites mais repose sur une simple relation de cooccurrence textuelle entre la mention cible et les mentions d'entités qui l'entourent. Nous faisons l'hypothèse que cette cooccurrence est la manifestation de relations entre les entités auxquelles réfèrent les mentions considérées.

L'exploitation conjointe des représentations des mentions, des entités et de la notion de contexte relationnel textuel dans le cadre défini par (Moreno *et al.*, 2017) nous a conduit à définir les quatre mesures de similarité suivantes entre une mention et une entité candidate :

- REL_1 la similarité directe entre le vecteur de la mention et de celui de l'entité candidate, mesurée par une distance cosinus ;
- REL_2 la similarité cosinus entre le vecteur moyen des mentions du contexte relationnel textuel et le vecteur de l'entité candidate ;
- REL_3 la valeur moyenne des similarités cosinus entre le vecteur de chaque mention du contexte relationnel textuel et le vecteur de l'entité candidate ;
- REL_4 la moyenne des k meilleures similarités parmi celles considérées pour REL_3 .

En notant \vec{m} le vecteur de la mention, \vec{e} le vecteur de l'entité et $p(m)$ le contexte relationnel de la mention, ces mesures se définissent de façon plus formelle par les équations suivantes :

$$REL_1(e, m) = \cos(\vec{e}, \vec{m})$$

$$REL_2(e, m) = \cos(\vec{e}, \frac{\sum_{w_i \in p(m)} \vec{w}_i}{||p(m)||})$$

$$REL_3(e, m) = \frac{\sum_{w_i \in p(m)} \cos(\vec{e}, \vec{w}_i)}{||p(m)||}$$

$$REL_4(e, m) = \frac{\sum \text{argmax}_{w_i \in p(m) | i=1, \dots, k} \cos(\vec{e}, \vec{w}_i)}{k}$$

2.3 Intégration dans un système général de désambiguïsation d'entités

Pour tester cette approche, nous intégrons le score généré par ces représentations dans un système global de désambiguïsation d'entités s'appuyant sur un apprentissage supervisé pour l'étape de désambiguïsation proprement dite. Ce système suit une architecture standard en deux grandes étapes :

- une étape de génération des entités candidates : à partir d'une mention d'entité, un ensemble d'entités candidates de la base de connaissances est généré en s'appuyant sur l'égalité ou la proximité de la mention avec le nom de l'entité ou un nom (Dredze *et al.*, 2010) ;
- une étape de sélection de la meilleure entité candidate, fondée sur un apprentissage supervisé. Cet apprentissage utilise les couples (mention, entité candidate) comme exemples positifs lorsque l'entité candidate est l'entité de référence et comme exemples négatifs sinon. Les exemples sont représentés par un vecteur de traits comprenant des traits indiquant la façon dont le candidat a été généré (*i.e.* le degré de proximité entre les chaînes de caractères de la mention et de l'entité candidate) ainsi que plusieurs mesures de similarité : une similarité cosinus entre les représentations vectorielles du document et de la description textuelle de l'entité et les similarités fondées sur les représentations jointes de mots et d'entités proposées par (Moreno *et al.*, 2017).

Ce système forme notre modèle de base (noté *baseline* dans les résultats). Le score de similarité relationnelle y est intégré par l'ajout d'un ou plusieurs traits supplémentaires.

3 Évaluation

3.1 Corpus et mesures

Pour l'évaluation de nos méthodes, nous utilisons les données de la tâche de désambiguïsation d'entités nommées (EDL, pour *Entity Discovery and Linking*) de la campagne TAC KBP de 2015². Nous présentons dans le tableau 2 des statistiques sur ces données. La base de connaissances de référence utilisée pour lever l'ambiguïté des entités est une sous-partie de la base Freebase (Bollacker *et al.*, 2008), contenant plus de 8 millions d'entités. Chaque document des collections d'entraînement (*train*) et de test contient un ensemble de mentions d'entités annotées avec les entités auxquelles elles font référence dans Freebase. Lorsqu'aucune entité de la base de connaissances ne correspond à la mention, celle-ci est annotée en tant que *NIL*.

TAC 2015	Nb. docs	Nb. mentions	Nb. mentions NIL	Nb. candidats	Nb candidats NIL
<i>train</i>	168	12 175	3 215	1 722 518	1 745
<i>test</i>	167	13 587	3 379	1 720 767	1 751

TABLE 2 – Description des données de la campagne TAC KBP 2015

2. <https://tac.nist.gov//2015/KBP/>

Pour chacune des mentions d’entraînement et de test, la génération des candidats telle que décrite à la section 2.3 permet d’obtenir un certain nombre d’entités candidates. Si pour une mention donnée, il n’a été trouvé aucun candidat potentiel, elle est classée comme *NIL*. Nous rapportons également le nombre de candidats et de *NIL* générés dans le tableau 2.

Pour évaluer les performances de notre modèle, nous utilisons les mesures de précision (P) et rappel (R) sur le candidat sélectionné par le système, s’il existe (*link*), ou sur les réponses *NIL* sinon. Ces mesures sont combinées de façon standard par un f-score (F). Si l’on note e_r , l’entité de référence associée à la mention m , e_t l’entité associée à m par notre système et $N(x)$, le nombre de mentions d’entités qui vérifiant l’expression x , les mesures utilisées se définissent par :

$$\begin{aligned}
 P(NIL) &= \frac{N(e_t=NIL \ \& \ e_r=NIL)}{N(e_t=NIL)} & P(link) &= \frac{N(e_t=e_r \ \& \ e_t \neq NIL)}{N(e_t \neq NIL)} & P(all) &= \frac{N(e_t=e_r)}{N(e_t)} \\
 R(NIL) &= \frac{N(e_t=NIL \ \& \ e_r=NIL)}{N(e_r=NIL)} & R(link) &= \frac{N(e_t=e_r \ \& \ e_t \neq NIL)}{N(e_r \neq NIL)}
 \end{aligned}$$

Notons que $P(all) = R(all) = F(all)$ si le classifieur a donné une réponse pour toutes les mentions ($N(e_t) = N(e_r)$).

Le système de classification supervisé utilisé pour la désambiguïsation s’appuie sur un modèle non déterministe (en particulier pour faire des échantillonnages aléatoires des données d’entraînement). Les scores présentés dans cette étude sont des moyennes des scores obtenus sur 10 tests. Pour avoir une comparaison correcte des différents modèles, chacun des tests est réalisé avec la même graine aléatoire pour les différents modèles.

3.2 Résultats

Les résultats présentés dans cette section ont été obtenus avec une taille du contexte relationnel des entités limité à 1 : on ne considère donc que des séquences de trois éléments pour la construction des représentations avec *word2vec*. Cette construction a été réalisée avec le modèle CBOW, une taille des représentations égale à 400 et une seule itération. Au niveau du système global de désambiguïsation, le classifieur utilisé pour la sélection de la meilleure entité candidate est Adaboost.

L’intégration dans le système des 4 similarités relationnelles (REL_1 , REL_2 , REL_3 et REL_4) a été testée de façon indépendante (ajout d’une des similarités comme trait pour le classifieur) ou combinée (ajout de plusieurs de ces similarités comme traits pour le classifieur). Nos expériences ont en effet montré que combiner les mesures peut apporter des améliorations.

	P(all)	P(NIL)	R(NIL)	F(NIL)	P(link)	R(link)	F(link)
baseline	0,765	0,624	0,908	0,74	0,845	0,718	0,776
REL_1	0,776	0,626	0,909	0,741	0,861	0,732	0,792
$REL_{1,2}$	0,781	0,626	0,92	0,745	0,871	0,735	0,797
$REL_{1,2,3,4}$	0,773	0,614	0,915	0,735	0,867	0,727	0,79
$REL_{1,2}$ + type d’entité	0,742	–	–	–	–	–	–
Top TAC-EDL 2015 (Moreno <i>et al.</i> , 2017)	0,737	–	–	–	–	–	–
	0,742	–	–	–	–	–	–

TABLE 3 – Résultats de désambiguïsation d’entités sur TAC-EDL 2015, sur une moyenne de 10 runs

Nous présentons dans le tableau 3 les résultats les plus significatifs. Dans tous les cas, nous observons que l'ajout d'une mesure de similarité utilisant les représentations des relations de la base de connaissances permet d'améliorer les résultats par rapport au modèle de base. Parmi les différentes mesures de similarité considérées de façon indépendante, la mesure directe REL_1 est la plus performante, avec une amélioration du score global de 1,1 points. Combiner les mesures de similarité permet d'améliorer encore les résultats mais conjuguer de façon naïve la totalité des mesures donne des résultats inférieurs à ceux obtenus avec la seule mesure REL_1 . La meilleure configuration associée en fait les mesures REL_1 et REL_2 , avec un gain de 1,6 points par rapport à la *baseline*. On peut ainsi supposer que combiner une mesure directe et une mesure prenant en compte un contexte un peu plus large de la mention est intéressant à condition toutefois que l'influence de ce contexte ne noie pas trop l'information apportée par la mesure directe.

Dans le détail, on constate que le gain porte surtout sur les scores de type *link* (précision et rappel), c'est-à-dire que le modèle aide à repérer les bonnes entités lorsqu'elles existent. Mais on observe également que l'amélioration sur le rappel pour les entités NIL augmente de façon notable. De façon générale, la différence sur les entités NIL entre le score élevé pour le rappel et un moins bon score pour la précision montre que l'approche définie tend à classer un peu trop d'entités comme NIL.

Nous comparons également nos résultats à l'état de l'art. Les mesures officielles de la campagne se fondent à la fois sur la détection de l'entité et sur la détection du type de l'entité (y compris pour les entités NIL). Les scores rapportés ci-dessus n'incluent pas cette notion de type. En ajoutant cette contrainte, notre meilleur modèle obtient un score de 0,742. Le meilleur participant à la campagne TAC EDL 2015 avait un score global de 0,737 et Moreno *et al.* (2017) rapportent également, sur cette collection, un score de 0,742. Les résultats obtenus ici sont donc comparables à l'état de l'art. On note que Moreno *et al.* (2017) utilisent également d'autres indices, telles que la popularité *a priori* d'une entité, qui ne sont pas exploités ici et que l'on pourrait ajouter à notre système comme traits supplémentaires.

4 Conclusion et perspectives

Nous présentons dans cet article une approche pour l'exploitation des relations entre les entités présentes dans une base de connaissances pour aider la désambiguïsation des entités nommées. Cette méthode s'appuie sur un apprentissage de représentations pour les relations (*relation embeddings*) exploitant le modèle *word2vec* sur les contextes relationnels des entités dans la base de connaissances. En évaluant ce modèle sur les données de la campagne d'évaluation TAC EDL 2015, nous montrons que l'ajout de cette connaissance permet d'améliorer les performances d'un système de désambiguïsation d'entités de façon significative. Comme nous l'avons montré, la prise en compte d'un certain contexte autour d'une mention d'entité est intéressante mais doit être contrôlée soigneusement. Une extension possible de ce travail est de considérer le contexte à un niveau plus large, en l'occurrence celui du document, en réalisant de façon jointe la désambiguïsation d'un ensemble de mentions à l'instar des travaux menés à la suite de (Cucerzan, 2007) ou de (Kulkarni *et al.*, 2009) par exemple.

Références

base de connaissances pour la désambiguïsation d'entités nommées. In *Actes de la 23e conférence sur le Traitement Automatique des Langues Naturelles*, p. 290–303, Paris, France: Association pour le Traitement Automatique des Langues.

BOLLACKER K., EVANS C., PARITOSH P., STURGE T. & TAYLOR J. (2008). Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, p. 1247–1250, Vancouver, Canada: ACM.

CUCERZAN S. (2007). Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2007)*, p. 708–716, Prague, Czech Republic.

DREDZE M., MCNAMEE P., RAO D., GERBER A. & FININ T. (2010). Entity Disambiguation for Knowledge Base Population. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10)*, p. 277–285, Beijing, China: Association for Computational Linguistics.

FANG W., ZHANG J., WANG D., CHEN Z. & LI M. (2016). Entity Disambiguation by Knowledge and Text Jointly Embedding. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL 2016)*, p. 260–269, Berlin, Germany.

KULKARNI S., SINGH A., RAMAKRISHNAN G. & CHAKRABARTI S. (2009). Collective Annotation of Wikipedia Entities in Web Text. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009)*, p. 457–466, Paris, France: ACM.

LEHMANN J., ISELE R., JAKOB M., JENTZSCH A., KONTOKOSTAS D., MENDES P., HELLMANN S., MORSEY M., VAN KLEEF P., AUER S. & BIZER C. (2015). DBpedia – A large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, **6**(2), 167–195.

LING X., SINGH S. & WELD D. (2015). Design Challenges for Entity Linking. *Transactions of the Association for Computational Linguistics (TACL)*, **3**, 315–328.

MIKOLOV T., SUTSKEVER I., CHEN K., CORRADO G. S. & DEAN J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26*, p. 3111–3119. Curran Associates, Inc.

MORENO J., BESANÇON R., BEAUMONT R., D'HONDT E., LIGOZAT A.-L., ROSSET S., TANNIER X. & GRAU B. (2017). Combining Word and Entity Embeddings for Entity Linking. In *The Semantic Web. ESWC 2017*, volume 10249 of *Lecture Notes in Computer Science*: Springer.

MORO A., RAGANATO A. & NAVIGLI R. (2014). Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics (TACL)*, **2**, 231–244.

NICKEL M., TRESP V. & KRIEGEL H.-P. (2012). Factorizing YAGO: scalable machine learning for linked data. In *Proceedings of the 21st international conference on World Wide Web (WWW 2012)*, p. 271–280, Lyon, France: ACM.

PEROZZI B., AL-RFOU R. & SKIENA S. (2014). DeepWalk: Online Learning of Social Representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2014)*, p. 701–710, New York, USA: ACM.

RIEDEL S., YAO L., MCCALLUM A. & MARLIN B. M. (2013). Relation Extraction with Matrix Factorization and Universal Schemas. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2013)*, p. 74–84, Atlanta, Georgia, USA.

- SHEN W., WANG J. & HAN J. (2015). Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions. *Transactions on Knowledge & Data Engineering*, **27**(2), 443–460.
- SUCHANEK F. M., KASNECI G. & WEIKUM G. (2007). Yago: A Core of Semantic Knowledge. In *Proceedings of the 16th International Conference on World Wide Web (WWW 2007)*, p. 697–706, Banff, Alberta, Canada: ACM.
- USBECK R., NGOMO A.-C. N., RÖDER M., GERBER D., COELHO S., AUER S. & BOTH A. (2014). AGDISTIS - Graph-Based Disambiguation of Named Entities using Linked Data. In *The Semantic Web – ISWC 2014*, volume 8796 of *Lecture Notes in Computer Science*, p. 457–471. Springer International Publishing.
- VLANDEČIĆ D. & KRÖTZSCH M. (2014). Wikidata: A Free Collaborative Knowledgebase. *Communications of the ACM*, **57**(10), 78–85.
- WESTON J., BORDES A., YAKHNENKO O. & USUNIER N. (2013). Connecting Language and Knowledge Bases with Embedding Models for Relation Extraction. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*, p. 1366–1371, Seattle, Washington, USA.
- YAMADA I., SHINDO H., TAKEDA H. & TAKEFUJI Y. (2016). Joint Learning of the Embedding of Words and Entities for Named Entity Disambiguation. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL 2016)*, p. 250–259, Berlin, Germany.

