

Savana: Un Entorno Integral de Extracción de Información y Expansión de Terminologías en el Dominio de la Medicina

Savana: A Global Information Extraction and Terminology Expansion Framework in the Medical Domain

Luis Espinosa-Anke*†, Jorge Tello†, Alberto Pardo**, Ignacio Medrano†
Alberto Ureña†, Ignacio Salcedo†, Horacio Saggion*

*TALN-DTIC, Universitat Pompeu Fabra, Carrer Tànger, 122 08018 -Barcelona

†Medsavana S.L., C/ Jiloca 4 - 5 Derecha, 28016 - Madrid

**Fundación Jiménez Díaz

{luis.espinosa,horacio.saggion}@upf.edu

{jtello,ihmedrano,aurena,isalcedo}@savanamed.com

alberto.pardo@quironsalud.es

Resumen: Las bases terminológicas médicas constituyen una fuente de información fundamental en el dominio médico, ya que son utilizadas a diario tanto por profesionales en el sector como en el ámbito académico. Existen numerosos recursos de este tipo, tales como la Clasificación Internacional de Enfermedades (CIE), SnomedCT, o UMLS (Unified Medical Language System). La calidad de estas bases terminológicas es en general alta, dado que están construidas manualmente por expertos. Sin embargo, su capacidad para representar fielmente un dominio como el médico, que se encuentra en constante evolución, es limitada. Por tanto, el desarrollo de sistemas capaces de capturar nuevo conocimiento en fuentes textuales heterogéneas e incluirlas en terminologías estándar tienen el potencial de añadir un gran valor añadido a dichas terminologías. Este artículo presenta, en primer lugar, SAVANA, un sistema de extracción de información biomédica que, combinado con validación por parte de profesionales médicos, es utilizado para popular la rama española de SnomedCT con nuevo conocimiento. En segundo lugar, describimos y evaluamos un sistema que, dado un término médico nuevo, le asigna su hiperónimo más probable, constituyendo así un facilitador en tareas de enriquecimiento y expansión de bases terminológicas médicas.

Palabras clave: Terminologías médicas, bases de conocimiento, snomed, word2vec, semántica, savana

Abstract: Terminological databases constitute a fundamental source of information in the medical domain. They are used daily both by practitioners in the area, as well as in academia. Several resources of this kind are available, e.g. CIE, SnomedCT or UMLS (Unified Medical Language System). These terminological databases are of high quality due to them being the result of collaborative expert knowledge. However, they may show certain drawbacks in terms of faithfully representing the ever-changing medical domain. Therefore, systems aimed at capturing novel terminological knowledge in heterogeneous text sources, and able to include them in standard terminologies have the potential to add great value to such repositories. This paper presents, first, SAVANA, a Biomedical Information Extraction system which, combined with a validation phase carried out by medical practitioners, is used to populate the Spanish branch of SnomedCT with novel knowledge. Second, we describe and evaluate a system which, given a novel medical term, finds its most likely hypernym, thus becoming an enabler in the task of terminological database enrichment and expansion.

Keywords: Medical terminologies, knowledge bases, snomed, word2vec, semantics, savana

1 Introduction

Today, in the Information Age, we are witnessing an unprecedented growth in the production and availability of knowledge stored in the web, a massive source of publications, source code, data, research websites, wikis and blogs (O’Donoghue et al., 2014). The Health domain is not oblivious to this (r)evolution, and both academics and practitioners are starting to leverage medical evidence obtained from large-scale knowledge repositories in order to enhance the so-called Evidence-Based Medicine (Kumar, 2011). One of the great challenges for enabling data-driven support to clinical decisions is making sense of unstructured information appearing in research papers, text books, social networks or, what is more relevant for this paper, clinical records.

Clinical Health Records (CHRs) are documents where doctors take notes on a patient’s medical condition, his or her progress, and suggest possible medication and treatment. They are a rich source of information because they provide personalized empirical data on treatment and evolution of medical conditions, and hence this type of document is receiving interest from the NLP community as enabler not only for medical support systems but also for its potential as training and evaluation data for Machine Learning algorithms in the field of Bioinformatics.

Examples of the interaction between NLP and CHRs include MEDEX (Xu et al., 2010), a system for extracting medication information from clinical narratives, or a system for drug reaction event extraction (Santiso et al., 2016). However, and despite their potential, CHRs pose great challenges for automatic processing, as they are often unstructured, ill-defined and arduous to analyse at scale (Iqbal et al., 2015).

In this context, Medical Terminological Databases (MTDs) play a crucial role, as they provide a structured ground where medical concepts and their relations are encoded by medical experts and can be used as a benchmark for developing algorithms that leverage medical concept extraction to some extent. One of the best known MTDs is SnomedCT (Spackman, Campbell, and Côté, 1997), which is part of the UMLS (Bodenreider, 2004). One of the main drawbacks of MTDs is that creating and maintaining them manually is arduous. More impor-

tantly, keeping them up to date is not possible considering the amount of novel information that is generated daily. Furthermore, even if they are manually created, there is certain discussion even on their quality, since it is difficult to control the fitness of every single addition to the database (Morrey et al., 2009).

In this paper we propose to bridge the gap between unstructured medical knowledge stored arbitrarily in CHRs, on one hand, and the automatic maintaining of MTDs, on the other. In the remainder of this paper, we first describe SAVANA, a Biomedical Information Extraction system, which we run on a large collection of CHRs. In a second phase, SAVANA’s predictions are presented to medical practitioners, who validate novel associations between SnomedCT concepts and their lexicalizations (i.e. the way they are expressed in free text). We exploit the combination of SAVANA and the validation stage to obtain a validation dataset of nearly 500 novel medical terms in Spanish, on which we evaluate several unsupervised systems aimed at finding, for each candidate novel term, its best point of attachment in the Spanish SnomedCT Database. These systems are based on both syntactic and semantic properties. Our results suggest that this is a promising direction for performing large-scale medical terminology extraction for Spanish, along with its *semantification*.

2 Related Work

The availability of MTDs is in constant growth. Paramount examples range from well-established collaborative efforts like UMLS (Bodenreider, 2004), which serves as an umbrella for multilingual resources such as SnomedCT (Spackman, Campbell, and Côté, 1997), or even the CIE database (*Clasificación Internacional de Enfermedades*), published by the Organización Panamericana de la Salud (1995). In addition, general purpose resources are increasingly playing more important roles in biomedical NLP tasks, as is the case of Wikipedia, which has been exploited for identifying medical disorders in CHRs (Bodnari et al., 2013).

Structured knowledge resources may present drawbacks such as staticity, and hence they may become obsolete in highly active and ever-changing domains such as the biomedical. Research in NLP has

proposed several approaches to alleviate this problem. For instance, in the area of learning lexical taxonomies, novel terms are discovered and included in domain-specific is-a hierarchies (Velardi, Faralli, and Navigli, 2013; Luu Anh, Kim, and Ng, 2014; Kozareva, 2014; Espinosa-Anke et al., 2016). Another approach is to combine Open Information Extraction paradigms with pre-defined semantic criteria. These criteria may be based on Wikipedia (Nakashole, Weikum, and Suchanek, 2012), BabelNet (Delli Bovi, Espinosa-Anke, and Navigli, 2015) or may be *ad-hoc* semantic hierarchies (Carlson et al., 2010). Finally, there are (fewer) approaches on discovering novel terminology from domain-specific dictionaries, glossaries or web pages and providing semantics via intersection with WordNet (Miller, 1995), e.g. by processing associated glosses or definitions (Jurgens and Pilehvar, 2015). Most of these approaches, however, while having shown notable success in the English language, did not address other languages, for which availability of resources and tools is scarcer.

The medical domain has also received attention in terms of automatically expanding existing resources. Prominent examples include (1) The development of novel MTDs from Wikipedia (Pedro, Niculescu, and Lita, 2008); (2) Enriching SnomedCT terminology with associated definitions (Ma and Distel, 2013); and in multilingual settings, (3) Expansion of SnomedCT in Swedish by processing CHRs (Henriksson et al., 2013).

3 Savana

We use SAVANA, a Biomedical Information Extraction System¹, integrated in several public and private healthcare institutions in Spain, for obtaining and validating ground truth data. The SAVANA algorithm is designed to retrieve prominent biomedical information from CHRs in the Spanish language. It does so by combining in its pipeline modules for, among others, sentence segmentation, tokenization, spell checking, acronym detection and expansion, negation identification, and a multi-dimensional ranking scheme which combines linguistic knowledge, statistical evidence, and state-of-the-art continuous vector representations of words and doc-

uments in the biomedical domain learned via shallow neural networks. We run SAVANA over several thousand CHRs, and ask medical practitioners to validate matches of SAVANA’s association between a mention of a medical concept in text, and an existing SnomedCT entry, by means of a web interface (Figure 1). The subset of the Spanish SnomedCT branch on which we run our experiments contains over 401,126 concepts, which are linked by means of 2,722,877 hypernymic (is-a) relations. The validation procedure may yield *novel terminology* in terms of either novel lexicalizations for an existing term (synonyms), or novel terms which can be attached to a more general SnomedCT concept (hyponyms). In this paper we are interested in the latter case: Finding the best point of attachment for novel concepts, rather than finding additional ways of expressing the same idea. At validation stage, if human experts consider that a concept identified by SAVANA has a meaning which is missing in SnomedCT, this concept makes it to our ground truth novel terminology, and hence will constitute the testbed for the experiments we describe in Section 4. We collect gold standard data of up to 492 novel terms, with an average of 3.2 hypernymic relations encoded by human experts. There was no restriction in the type of concept to be included. Therefore, this dataset includes diverse terms which are related to infrastructure, e.g. SERVICIO DE ODONTOLOGÍA → {servicio hospitalario}², or actual medical conditions, e.g. GONARTROSIS → {trastorno de la rodilla, enfermedad de la rodilla}. In the following section, we describe the experiments carried out to discover the most appropriate hypernym for each of the 492 novel terms we incorporated to SnomedCT thanks to combining the SAVANA algorithm with an expert validation stage.

4 Enriching SnomedCT

In this section we describe the SnomedCT enrichment experiments. Given a novel term, we aim at finding its best point of attachment, expressed as its closest hypernym. Our approach is unsupervised and hence requires no prior annotation or training. Moreover, we do not exploit any web or Wikipedia-based textual evidence (which we may inves-

¹<http://www.savamed.com/>

²We denote is-a relations between terms and sets of hypernyms as $term \rightarrow \{hypernym1, hypernym2\}$.

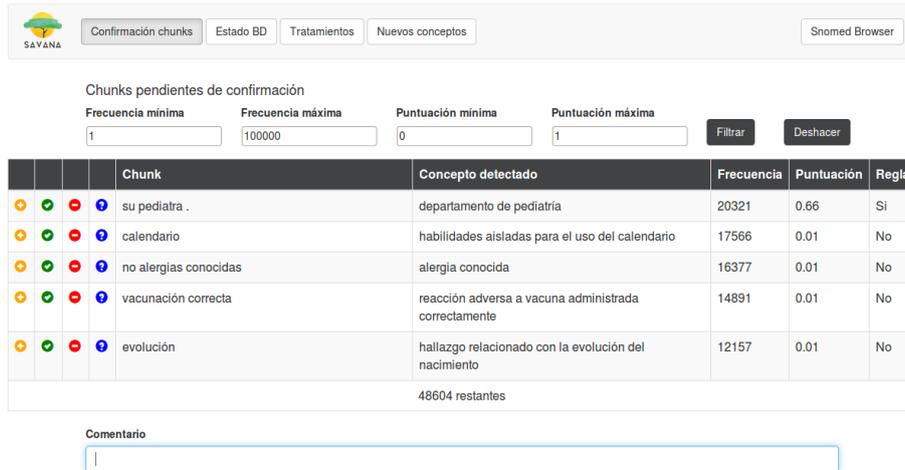


Figure 1: A snapshot of the validation web interface. Let us highlight how the validation procedure allows the medical expert to assign to the SnomedCT concept `departamento de pediatría`, a novel lexicalization in the context of CHRs, namely the string `su pediatra`.

tigate in future work). However, we do leverage two main resources in our experiments, which are described briefly.

- For syntactic parsing, we use a transition-based parser based on the parsing technology included in the Mate framework (Bohnet, 2010).
- For computing similarities between concepts, we exploit word embeddings derived from training a shallow neural net model (Mikolov, Yih, and Zweig, 2013) with the *word2vec*³ tool, implemented in *gensim*⁴. The model used for our experiments comes from a 2015 dump of the Spanish Wikipedia preprocessed and lemmatized with Freeling (Atserias et al., 2006). Our model is 300-dimensional, and is trained using the skip-gram with negative sampling algorithm, using a minimum count of 10 for each word.

Having described the two main technological pivots of our approach, let us describe each of the systems evaluated:

- **Substring**⁵ This is a substring inclusion baseline which, for each novel term, assigns as term hypernyms all Snomed concepts that are subsumed in the novel

term. For example, given the unseen concept GONARTROSIS, candidate hypernyms are ARTROSIS and ARTROSIS (TRASTORNO). Note that this approach fails short when dealing with longer and more complex terminology, as in the case of the concept NO OTROS HÁBITOS TÓXICOS, where incorrect hypernyms are captured, such as TOS or OTRO.

- **Head Fuzzy Match*** Multiword terms (mwt) may be generalized via their syntactic dependencies. For example, given the novel concept INSUFICIENCIA CARDÍACA CONGESTIVA LEVE, after dependency parsing we are able to isolate INSUFICIENCIA as the mwt’s head. This configuration of our approach collects all Snomed concepts of which this head is substring. In this example, we would correctly match INSUFICIENCIA CARDÍACA, but also generate false positives such as INSUFICIENCIA HEPÁTICA or INSUFICIENCIA RESPIRATORIA TIPO 2.
- **Head Exact Match** This is a restricted version of *Head Fuzzy Match*, in which in most cases we only obtain one candidate, i.e. the Snomed concept which matches exactly the out-of-vocabulary (OOV) term’s head. For instance, for the concept NO OTROS HÁBITOS TÓXICOS, the retrieved candidate would be the Snomed concept

³code.google.com/archive/p/word2vec/

⁴radimrehurek.com/gensim/models/word2vec.html

⁵We distinguish baseline systems with *.

HÁBITO.

- **Distributional** The first of our distributional approaches, exploiting word embeddings, stems from the intuition that similar concepts may occur in similar contexts. This property has been confirmed to hold in many semantic relations (Mikolov, Yih, and Zweig, 2013; Mikolov et al., 2013). In this configuration of our system, given a term t consisting of a set of words $\{w_i, \dots, w_n\}$ (after stopword removal), we compute the centroid vector μ of the set of associated word vectors $\vec{w} \in t$. We obtain $\mu(t)$ as follows:

$$\mu(t) = \frac{1}{|t|} \sum_{\vec{w} \in t} \frac{\vec{w}}{\|\vec{w}\|} \quad (1)$$

We perform the same operation on all candidate Snomed concepts. Specifically, we obtain, given a Snomed terminology \mathcal{S} , for each Snomed term $t_s \in \mathcal{S}$, its corresponding centroid vector $\mu(t_s)$. Then, our algorithm returns as best match the Snomed concept maximizing the semantic similarity between t and t_s , denoted as $\text{SIM}(t, t_s)$, and computed via cosine score as follows:

$$\text{SIM}(t, t_s) = \frac{\mu(t) \cdot \mu(t_s)}{\|\mu(t)\| \|\mu(t_s)\|} \quad (2)$$

This operation yields a ranked list of candidates by score, where score is the cosine score above, and thus the predicted candidate is the term t_s with the highest similarity with the input term t .

- **DistDep** Our last experiment is performed with the **DistDep** system, which combines head word lookup with similarities derived from word embeddings. It simply consists in comparing the vector associated to the head node (as extracted in any of the **Head**-based approaches) of the novel term with the centroid of all available concepts in \mathcal{S} , and keeping as best match the highest scoring candidate.

5 Evaluation

5.1 Distance-based Evaluation

The evaluation of a system’s performance in terms of its ability to attach novel ter-

minology to an existing knowledge repository is traditionally performed by considering distance between reference nodes and predicted nodes, i.e. the best point of attachment, and the decision made by the system. While evaluation metrics exist for computing semantic similarity over lexical databases like WordNet, these are not suitable in our case because our branch of Snomed is designed in a slightly different fashion, as it can be considered a multiroot directed acyclic graph (DAG), and hence in many cases, given two concepts, there is no *least common subsumer* other than the root node. For instance, the path between TRASTORNO CON TALLA BAJA and PRUEBA DE VARIANTE DE HEMOGLOBINA includes one of the root nodes in the SnomedCT taxonomy, namely SNOMED CLINICAL TERMS (ENERO 2014).

This makes metrics like Wu&Palmer Similarity (Wu and Palmer, 1994), which considers lowest common subsumers in their similarity computation, unsuitable. For this reason, we propose a distance metric for evaluation purposes sensitive to terminological databases shaped as DAGs rather than trees (like WordNet). We account for the fact that there may be several valid points of attachment and hence compute an average of node-based shortest path $sp(\cdot)$ over all predicted candidates and all gold standard nodes.

Let \mathcal{G} be the set of gold standard points of attachment to a novel term t , and \mathcal{P} the set of predictions generated by a system. We define an Error-Score function E that, given a novel term t , computes the average shortest path of all predicted points of attachment $p \in \mathcal{P}$:

$$E(t) = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \frac{\sum_{g \in \mathcal{G}} sp(p, g)}{|\mathcal{G}|} \quad (3)$$

In addition, we report a second evaluation based on whether a system is able to capture all the gold standard points of attachment, regardless of additional incorrect predictions. This is performed only on those systems which return a *set of candidates*, which is not the case of the distributional systems **Distributional** and **DistDep**. We included evaluation under this Recall score, which we denote as $R(t)$, as we foresee a real world scenario where human post-edition of false positives may be less time-consuming than finding in SnomedCT the best points of attachment for each novel term. We simply

set $R(t) = 1$ if a given approach is able to cover, with its n predictions, all the possible gold standard attachments, and $R(t) = 0$ otherwise, and average results over the total prediction sets. We provide the evaluation results for both criteria in Table 1. The two main conclusions that can be drawn from our experimental results are that, first, leveraging similarities derived from word embeddings improve the performance of MTD enrichment systems, and second, that exploiting a greedy approach of fuzzy syntactic head matching is a reasonable strategy for increasing recall.

	Error-Score	Recall
Substring*	8.51	26%
Head Fuzzy	7.07	84%
Head Exact	4.72	13%
Distributional	3.34	N/A
DistDep	3.36	N/A

Table 1: Evaluation results for our proposed systems in terms of average performance of all its predictions (Error-Score), and Recall.

Finally, we plotted the performance in Error-Score of our three proposed systems (not baselines) over all the novel terms present in the evaluation data. We can observe that the two distributional systems based on word embeddings show a similar behaviour, much better in general than the third best system, **Head Exact** (Figure 2).

5.2 Human Evaluation

We assume in our automatic evaluation that human experts in the biomedical domain will provide a solid ground truth against which system predictions can be evaluated. However, given the size of SnomedCT, our system may provide correct points of attachment for novel terminology which were not included in the first place, and this is penalized in the automatic evaluation. For this reason, we presented human experts with the set difference between the sets of gold and predicted points of attachment, and asked them to label them as correct or incorrect. We find an average of 27% correctness over all systems, which suggests that certain cases of *false positives* were actually correct predictions and hence were valid inclusions of novel terms along

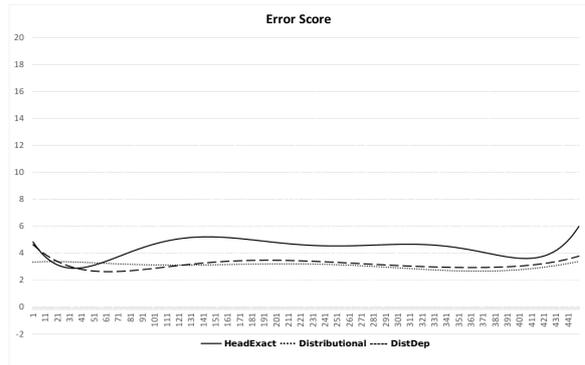


Figure 2: Error-Score results (y axis) for the three systems we propose, namely **Distributional** (dotted), **DistDep** (dashed), and **Head Exact Match** (line), over the whole test terminology (x axis).

with their associated hypernymic relations. We illustrate a few cases of *false positives* in Table 2 together with their correctness according to a domain expert.

6 Conclusions and Discussion

The rapidly growing interplay between Artificial Intelligence and healthcare is producing innovative assistive technologies (e.g. adaptive and rehabilitative devices), as well as medical support systems which leverage large quantities of heterogeneous data. Among the latter, let us highlight SAVANAMED, which thanks to the SAVANA algorithm, provides a real-time medical support system by making sense of textual information present in CHRs. In this paper, we described how the SAVANA algorithm, backed up by a validation stage carried out by medical practitioners, was used to produce a ground truth for evaluating a system in the task of MTD enrichment. We evaluated several systems against this data and found that combining linguistic information derived from syntactic dependencies, as well as similarities computed over word produces the best results. To the best of our knowledge, both SAVANA and the MTD enrichment system are the first systems of their kind developed for the Spanish language.

Our encouraging results open up a promising line of research in NLP tasks like semantics and Information Extraction in the medical domain, both industrial and academic

Novel Term	\mathcal{G}	Novel PoA (fp)	Correctness
dermatitis seborreica leve	eccema seborreico	dermatitis	Yes
servicio de cardiología pediátrica	servicio hospitalario	servicio de cardiología	Yes
artrosis cervical	artrosis	linfadenopatía cervical	No
talla baja idiopática	trastorno con baja estatura	al examen: estatura baja	No

Table 2: Illustrative cases where some of the novel concepts discovered by our approach, and evaluated as false positive (fp) by the automatic criteria, were considered correct in a second pass by human domain experts.

purposes, where we expect that close collaboration may result in larger and better datasets both for AI and NLP, and also for medical practitioners.

7 Future Work

We have presented an unsupervised approach for enriching medical terminological databases in Spanish. Our original idea was to explore to what extent we could *do well* in a setting without external knowledge being included in the pipeline (web, Wikipedia, and so forth), mainly having scalability in mind. However, we are interested in expanding our initial approach with this external knowledge, in the hope that it will contribute to discover additional candidates a medical terms, and as evidence for finding their best point of attachment.

We would also like to expand our experiments by including word embeddings coming from neural nets training, but trained on medical corpora. We are currently gathering data both from known medical websites, as well as a pool of medical records. We would also like to expand the system described in this paper with knowledge mined from scientific papers in order to also take advantage of information expressed in more canonical and less noisy fashion.

Acknowledgements

We would like to thank the three anonymous reviewers for their helpful comments. This work is partially funded by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502). The authors also acknowledge support from SAVANA and Hospital Universitario Fundación Jiménez Díaz.

References

- Atserias, Jordi, Bernardino Casas, Elisabet Comelles, Meritxell González, Lluís Padró, and Muntsa Padró. 2006. FreeLing 1.3: Syntactic and Semantic Services in an Open-Source NLP Library. In *Proceedings of LREC*, volume 6, pages 48–55.
- Bodenreider, Olivier. 2004. The Unified Medical Language System (UMLS): Integrating Biomedical Terminology. *Nucleic acids research*, 32(suppl 1):D267–D270.
- Bodnari, Andreea, Louise Deleger, Thomas Lavergne, Aurelie Neveol, and Pierre Zweigenbaum. 2013. A Supervised Named-Entity Extraction System for Medical Text. In *CLEF (Working Notes)*.
- Bohnet, Bernd. 2010. Very High Accuracy and Fast Dependency Parsing is Not a Contradiction. In *COLING*, pages 89–97.
- Carlson, Andrew, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. 2010. Toward an Architecture for Never-Ending Language Learning. In *Proceedings of AAAI*, pages 1306–1313.
- Delli Bovi, Claudio, Luis Espinosa-Anke, and Roberto Navigli. 2015. Knowledge Base Unification via Sense Embeddings and Disambiguation. In *Proceedings of EMNLP*, pages 726–736.
- Espinosa-Anke, Luis, Horacio Saggion, Francesco Ronzano, and Roberto Navigli. 2016. ExTaSem! Extending, Taxonomizing and Semantifying Domain Terminologies. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16)*.
- Henriksson, Aron, Maria Skeppstedt, Maria Kvist, Martin Duneld, and Mike Conway. 2013. Corpus-driven Terminology Development: Populating Swedish

- SNOMED CT with Synonyms Extracted from Electronic Health Records. *ACL 2013*, page 36.
- Iqbal, Ehtesham, Robbie Mallah, Richard George Jackson, Michael Ball, Zina M Ibrahim, Matthew Broadbent, Olubanke Dzahini, Robert Stewart, Caroline Johnston, and Richard JB Dobson. 2015. Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register. *PloS one*, 10(8):e0134208.
- Jurgens, David and Mohammad Taher Pilehvar. 2015. Reserating the Awesometastic: An Automatic Extension of the WordNet Taxonomy for Novel Terms. In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies, Denver, CO*, pages 1459–1465.
- Kozareva, Zornitsa. 2014. Simple, Fast and Accurate Taxonomy Learning. In *Text Mining*. Springer, pages 41–62.
- Kumar, Dhavendra. 2011. The Personalised Medicine: A Paradigm of Evidence-based Medicine. *Annali dell’Istituto superiore di sanit *, 47(1):31–40.
- Luu Anh, Tuan, Jung-jae Kim, and See Kiong Ng. 2014. Taxonomy Construction Using Syntactic Contextual Evidence. In *EMNLP*, pages 810–819, October.
- Ma, Yue and Felix Distel. 2013. Learning Formal Definitions for SNOMED CT From Text. In *Artificial Intelligence in Medicine*. Springer, pages 73–77.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.
- Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic Regularities in Continuous Space Word Representations. In *HLT-NAACL*, pages 746–751.
- Miller, George A. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41.
- Morrey, C Paul, James Geller, Michael Halper, and Yehoshua Perl. 2009. The Neighborhood Auditing Tool: A Hybrid Interface for Auditing the UMLS. *Journal of biomedical informatics*, 42(3):468–489.
- Nakashole, Ndapandula, Gerhard Weikum, and Fabian M. Suchanek. 2012. PATTY: A Taxonomy of Relational Patterns with Semantic Types. In *Proceedings of EMNLP-CoNLL*, pages 1135–1145.
- O’Donoghue, DP., H Saggion, F Dong, D Hurley, Y Abgaz, X Zheng, O Corcho, JJ Zhang, J-M Careil, B Mahdian, et al. 2014. Towards Dr Inventor: A Tool for Promoting Scientific Creativity. In *ICCC*.
- Organizaci3n Panamericana de la Salud. 1995. Clasificaci3n estadística internacional de enfermedades y problemas relacionados con la salud: d cima revisi3n: CIE-10.
- Pedro, V, S Niculescu, and L Lita. 2008. Okinet: Automatic extraction of a Medical Ontology from Wikipedia. In *WiKIAI08: a workshop of AAAI2008*.
- Santiso, Sara, Arantza Casillas, Alicia P rez, Maite Oronoz, and Koldo Gojenola. 2016. Document-level Adverse Drug Reaction Event Extraction on Electronic Health Records in Spanish. *Procesamiento del Lenguaje Natural*, 56:49–56.
- Spackman, Kent A, Keith E Campbell, and Roger A C t . 1997. SNOMED RT: A Reference Terminology for Health Care. In *Proceedings of the AMIA annual fall symposium*, page 640. American Medical Informatics Association.
- Velardi, Paola, Stefano Faralli, and Roberto Navigli. 2013. Ontolearn Reloaded: A Graph-based Algorithm for Taxonomy Induction. *Computational Linguistics*, 39(3):665–707.
- Wu, Zhibiao and Martha Palmer. 1994. Verbs Semantics and Lexical Selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 133–138. Association for Computational Linguistics.
- Xu, Hua, Shane P Stenner, Son Doan, Kevin B Johnson, Lemuel R Waitman, and Joshua C Denny. 2010. MedEx: A Medication Information Extraction System for Clinical Narratives. *Journal of the American Medical Informatics Association*, 17(1):19–24.