Using Domain Lexicon and Grammar for Ontology Matching

Citation for published version:
Quesada Real, F, Bella, G, Mcneill, F & Bundy, A 2020, Using Domain Lexicon and Grammar for Ontology Matching. in *Proceedings of the Fifteenth International Workshop on Ontology Matching.* CEUR-WS.org, The Fifteenth International Workshop on Ontology Matching, Athens, Greece, 2/11/20.

Using Domain Lexicon and Grammar for Ontology Matching

Francisco José Quesada Real[1,2], Gábor Bella[3], Fiona McNeill[1], and Alan Bundy[1]

[1] University of Edinburgh, UK
{s1580097,fmcneill,a.bundy}@ed.ac.uk
[2] University of Cádiz, Spain
franciscojose.quesada@uca.es
[3] University of Trento, Italy
gabor.bella@unitn.it

Abstract. There are multiple ontology matching approaches that use domain-specific background knowledge to match labels in domain ontologies or classifications. However, they tend to rely on lexical knowledge and do not consider the specificities of domain grammar. In this paper, we demonstrate the usefulness of both lexical and grammatical linguistic domain knowledge for ontology matching through examples from multiple domains. We also provide an evaluation of the impact of such knowledge on a real-world problem of matching classifications of mental illnesses from the health domain. Our experimentation with two matcher tools that use very different matching mechanisms—LogMap and SMATCH—shows that both lexical and grammatical knowledge improve matching results.

Keywords: Ontology Matching · Domain-Knowledge · Domain Language · Domain Lexicon · Domain Grammar

1 Introduction

Ontology Matching (OM) aims at finding correspondences between the classes and instances of multiple ontologies [10]. Thus, OM processes are commonly carried out to solve heterogeneity problems that occur when multiple knowledge resources need to be integrated or used together. Among common approaches used in OM, the comparison of node labels has been one of the most performant and widely used techniques. While label matching has been addressed by the earliest matchers through simple methods such as string similarity, more complex cases such as syonymy, cross-lingual, or domain-specific matching need linguistically better-founded solutions [3]. The problem of matching domain ontologies or classifications is special because labels tend to mix elements of the general language with domain terms, and sometimes even grammatical forms that are domain-specific. In cross-domain matching scenarios, phenomena of meaning shifts, polysemy, and synonymy make the matching task even harder, such as in the emergency response domain where subdomains of police, healthcare, fire brigades, etc., need to be aligned [17, 18]. Another example is that of mapping standard classifications within the healthcare domain that, despite relying on precise domain terminology, express the same concepts in different ways, such as *'rupture of aorta'* versus *'aortic aneurysm, ruptured'*. Establishing precise mappings across standards has a major importance for cross-border health applications as they enable automated data integration methods [6].

A large number of matchers analyse natural language labels, on different levels of complexity. A common approach is to incorporate linguistic *background knowledge* (BK) into the matcher [9, 10].

SMATCH [14] relies on domain-independent BK: it uses WordNet [12] as an English domain-independent lexical database, and analyses labels using general grammatical tools such as tokeniser, lemmatiser, and syntactic parser. Other matchers, such as LogMap [15] or YAM-BIO [2], have been customised to integrate domain terminology to address specific matching challenges, such as biomedical terms. This results in increased performance on domain-specific matching; however, the longer the labels become, the more likely their grammatical structures and their use of general language become important, which cannot be covered by terminological knowledge alone. For this reason, some matchers, such as AML [11] or ALIN [8], integrate both domain-independent and domain-specific knowledge (e.g. WordNet together with biomedical resources).

All of these matchers, however, are limited to using lexical BK. While some of them do address grammar through basic domain-independent methods (tokenisation, lemmatisation, stop word elimination), they do not cope with cases where the grammar depends on the domain.

In this paper, we investigate the impact of both domain lexicon and domain grammar in ontology matching, mainly focusing on label-based matching.

The paper is organised as follows. In Section 2 we describe how domain knowledge appears in ontology labels. Section 3 focusses on different approaches that matchers may use to take advantage of domain knowledge. A case study on the health domain is presented in section 4, being evaluated in section 5. The paper finishes with some concluding remarks and future works included in section 6.

2 Domain Language in Ontology Labels

The use of specialised linguistic constructs is common in most domains of knowledge. The most obvious case is the use of specialised terminology, consisting of words and expressions that either are used exclusively within the context of a domain (such as *to deglaze* in cooking meaning *'to loosen bits of food which stuck on the bottom of a pan by adding liquid'*), or that gain a new meaning within a domain (such as *to clarify* which in cooking refers specifically to butter).

Domain-specific meaning can, however, also be vehicled by non-lexical means, a phenomenon that we globally call *domain grammar*. Domain grammar can be found even within the short labels typical of ontologies and classifications. Below we provide examples of domain language from specialised text, including ontology labels.

Domain terms. The *UK Civil & Protection Lexicon* (UKCP) defines the term *medevac* that means *medical evacuation*, itself considered a specialised term. In order to align these two terms, a matcher would either need lexical background knowledge that states their synonymy, or—in this specific case—word-level analysis in order to detect that one term is the abbreviated form of the other.

Domain acronyms. The acronym *REM* has many meanings; in the domain of neurology it means *rapid eye movement*. Again, in a matching task the acronym can be matched either through the use of domain lexical knowledge or through acronym detection.

Word derivation. Derivation rules allow the creation of words through the use of affixes, such as *voyeur* $\mapsto$ *voyeurism* or *anorexia* $\mapsto$ *anorexic*. While, as in these cases, domain language often relies on the derivational rules of general grammar, domain-specific derivational affixes and rules also exist, such as *candida* $\mapsto$ *candidiasis* in the medical domain. Even though the common approach in lexicography is to enumerate derived words as separate lexical entries, lexicons are often incomplete in practice due to the high productivity of affixes. Thus, grammar-based approaches to detecting the relatedness of derived terms can be useful, as when matching the label *fetishism* with *fetishistic disorder*.

Word inflection. Inflection rules are defined by general language; yet, particular inflected forms can be more or less specific to domains. A well-known example are cooking recipes where sentences tend to begin with verbs either in infinitive or imperative form (e.g. *'Peel the onions'*, which in French may be expressed either as *'Peler les ognons'* or as *'Pelez les ognons'*).

Specific uses of punctuation. In labels of the International Classification of Diseases (ICD), such as *'Hallucinogen use, unspecified with hallucinogen persisting perception disorder (flashbacks)'*, parentheses are used to provide clues for the interpretation of the label. Square brackets, commas, or parentheses are also widely used in ontologies, classifications, and data schemas, such as to provide units of measure for numerical values: *speed (km/h)*. The precise interpretation (e.g. relevance or not with respect to the matching task) of such punctuation and the text they delimit depends on the domain and the particular application at hand.

Domain syntax. The same *'Hallucinogen use...'* example from above shows that labels can use non-standard syntax. This is sometimes motivated by the context of use, such as the need to sort the labels alphabetically motivates the use of the adjective *unspecified* in a postpositive form. The phrase *hallucinogen persisting* *perception disorder*, on the other hand, includes syntax that is not considered as standard in general language but is common in medical text. While syntax may play a minor role in matching very short labels, for longer classification entries it may be taken into account by the matcher tool, as in the case of SMATCH that performs syntactic parsing.

3 Leveraging Domain Language for Ontology Matching

The hypothesis verified in this paper is that "*matching performance can be improved by relying on knowledge that is specific to domain language*". However, as domain language also incorporates elements of general language, our study also considers this aspect. Accordingly, we classify linguistic background knowledge with respect to being general or domain-specific, as well as with respect to being lexical or grammatical. This delineates the following four categories of knowledge: (1) general lexicon; (2) general grammar; (3) domain lexicon; and (4) domain grammar. Furthermore, we consider three different forms of grammatical knowledge with respect to the linguistic elements to which they apply: (a) phrase-level (syntax, dealing with the way words are organised within labels); (b) word-level (morphology, i.e. grammar that deals with the structure of words); and (c) character-level (e.g. orthography and use of punctuation). Due to the shortness of ontology and classification labels, we deem it sufficient to consider only these three levels of granularity of grammar.

General Lexicon A domain-independent resource that is commonly used is Princeton WordNet [12] which is a lexical database in which nouns, verbs, adjectives and adverbs are grouped into sets of synonyms, each expressing a different concept. All sets are semantically related between them with an *is_a* relationship, forming a taxonomy, in which the more general elements are at the top and the more specific are at the bottom levels.

Domain Lexicon There are multiple domain-specific resources such as lexicons or domain terminologies that contain the technical terms of an specific domain. In the literature, we can find different approaches to integrating these resources within WordNet [1, 17]. Their main goal is to append specialised knowledge to general knowledge currently represented in WordNet (e.g. *coronavirus* as a specialised type of *infection*). However there are cases in which the current representation of a word in WordNet differs from its meaning in the domain-specific resource (e.g. *evacuation* in WordNet and in the UKCP). In these cases, the integration is more complex and needs to be done in a supervised way [17].

The main advantage of using domain lexical knowledge is that matchers have an enriched BK and are able to find mappings of labels that include some of the added new terms. Moreover, when matching ontologies from multiple or partially different domains (such as reference health knowledge involving subdomains of healthcare), domain information can be leveraged for word sense disambiguation within the matching process, resulting in improved precision [5].

General Grammar Most matchers consider the grammar within labels for the matching process. In this case, they carry out some of the following tasks with independence from the domain of the resources to be matched [10].

– *Phrase Level Grammar*.
  - *Tokenisation*. Labels are segmented into tokens (e.g. "*medium-scale evacuation*" becomes <*medium, scale, evacuation*>).
  - *Acronym extraction*. Characters of tokens are used to extract/discover acronyms (e.g. "*Non Governmental Organisation*" becomes "*NGO*").
  - *String similarity*. Compare string labels considering different measures and return a value according to their similarity degree (e.g. "*Level of emergency*" and "*Level 1 emergency*" have a high similarity degree).
  - *Stopword elimination*. Tokens that are recognised as articles, prepositions, conjunctions are removed (e.g. "*level of emergency*" becomes "*level emergency*").

– *Word Level Grammar*.
  - *Lemmatisation*. Tokens are reduced to basic forms (e.g. "*disasters*" becomes "*disaster*").

– *Character Level Grammar*.
  - *Normalisation*. This task includes several subtasks such as: case normalisation, diacritics suppression, blank normalisation, digit suppression or punctuation elimination.

Domain Grammar There are cases in which applying the previous domain-independent tasks to domain-specific resources is counter-productive. For example, if we apply digit suppression and stopword elimination to the following labels: "*Level of emergency*", "*Level 1 emergency*", "*Level 2 emergency*", "*Level 3 emergency*"; the matcher might output that all labels represent the same knowledge. Another example appears when the case normalisation task is just limited to transform all characters within the label into lower case letters. In this case, if the label contains Roman numerals they might pass unnoticed after the case normalisation. For these reasons, it is necessary to consider domain-specific grammar and address it conscientiously. Below there are described the approaches that we have implemented in our research:

– *Phrase Level Grammar*. Finding clues or postscripts that recurrently appear within the labels in a domain is not unusual. In this case, it is necessary to analyse if they add enough knowledge to keep them in the label or it is worth suppressing them (e.g. "Mild cognitive impairment~~, so stated~~").

– *Word Level Grammar*. Implementing derivational morphology rules to transform a term from one part-of-speech into another is interesting because enriching matchers' BK with these words allows those matchers that do not mainly base the matching process on string similarity measures to discover new mappings. Domain words produced by derivational morphology are added to matchers' BK as related forms (e.g. "*pathological*" is added as a related form of "*pathology*).

– *Character Level Grammar*. Depending on the domain, particularly in application domain knowledge resources, orthography follows different conventions. This makes necessary to address it optimally in each case. For example, there might be cases in which the content within parentheses or square brackets is meta-information that is not relevant for the meaning of the label (e.g. "*Post(-) traumatic stress disorder*"), being recommendable its suppression, whereas in other cases this content might be essential (e.g. "*Stable lodine (Potassium lodate tables)*").

The rules of the different domain grammar levels can be extracted both in a supervised or unsupervised way. The latter requires a huge number of documents to apply statistical methods, whereas the former does not need such quantity of documents, but involves more effort. In general, the rules at the *word level* can be transferred to any ontology within a domain (e.g. health), while the rules at the *phrase and character levels* usually are more dependent on the application domain (e.g. Hospitals of North London).

4 Case Study on the Health Domain

The main motivation lies in the need of solving semantic interoperability problems within the health domain. For example, when clinicians have to exchange health records that contain descriptions from multiple official classifications of diseases. To do so, we have developed several extensions to enrich the matcher's BK with health lexical and grammatical knowledge.

Due to descriptions of disorders containing not only technical, but also general terms, WordNet has been used as a domain-independent BK into which the extensions are plugged. The extensions have been developed following the Lexical Markup Framework (LMF) standard [13], and integrated into WordNet using Diversicon [4], which is a framework that allows extending WordNet with any domain-specific knowledge represented in LMF, validating and generating an enriched WordNet.

General Lexicon Princeton WordNet has been used as domain-independent resource. The main reason is that it represents general knowledge and there are multiple approaches that we could apply to enrich WordNet with domain-specific knowledge resources.

Health Domain Lexicon We have developed an extension for WordNet that includes health lexical knowledge extracted from the following resources:

– *MeSH* is the National Library of Medicine's controlled vocabulary thesaurus [16]. It consists of sets of terms, naming descriptions, in a hierarchical structure that permits searching at various levels of specificity. The hierarchy is sorted considering several semantic relations such as *is_a* or *part_of*. This hierarchy is similar to the way in which WordNet is organised, which makes easier its integration. The developed extension for WordNet contains all descriptions included in the "*Diseases*" and "*Psychiatry and Psychology*" MeSH categories. In this case, we only consider the *is_a* semantic relation, because we have detected several problems using *part_of* when matching diseases (e.g. a "*heel disease*" is a "*foot disease*", but an "*eye disease*" is not a "*face disease*"). Addressing these problems is something that we are considering as a future work.

- The *SPECIALIST* lexicon is an English lexicon which contains both commonly occurring English words and biomedical vocabulary [7]. It is composed of lexical records, being each of them formed by a base form and a set of spelling variants or morphological derivations. For example, the lexical entry with base "*nephroprotective*" (adj) has as spelling variant: "*nephro-protective*", and as morphological derivation "*nephroprotectivity*" (noun). This resource has been used for enriching matchers' BK lexically, through developing an extension for WordNet that contains all lexical entries included in SPECIALIST.

General Grammar It has been addressed applying the grammatical techniques included in the matchers by default and including general derivational morphology.

Phrase level Grammar. The tasks applied have been: tokenisation, string similarity and stop word elimination.

Word level Grammar. In this case we applied lemmatisation and the integration of general derivational morphology rules included in SPECIALIST. Table 1 shows examples of these rules.

Table 1. General derivational morphology rules.

Derivational rule	**Example**
iency$(noun) → ient$(adj)	immuno-deficiency(noun) → immuno-deficient(adj)
sation$(noun) → zed$(adj)	anesthetisation(noun) → anesthetized(adj)
ical$(adj) → y$(noun)	uroradiological(adj) → uroradiology(noun)
ism$(noun) → istic$(adj)	fetichism(noun) → fetichistic(adj)

Character level Grammar. The tasks applied have been case normalisation, blank normalisation and diacritics suppression.

Health Domain Grammar It has been addressed using health derivational morphology extracted from SPECIALIST, and considerations identified at *phrase* and *character grammar levels*. The former was used to enrich matchers' BK, whereas the latter were considered as a preprocessing step prior to the OM process.

Phrase level Grammar. In medical resources there are clues that recurrently appear within descriptions of disorders. Examples are "*, undefined*" and "*, so stated*". This meta-information does not add special value to labels, particularly affecting to those matchers that mainly use string similarity measures. The main reason is that they are penalised by irrelevant characters, which results in a lower similarity degree. Considering the previous issue we decided to suppress these interpretational clues from descriptions of diseases in a preprocessing step prior to the matching process.

Word level Grammar. Several domain-specific derivational morphology rules have been extracted from the SPECIALIST lexicon and integrated into WordNet. Examples of these rules are shown in table 2.

Table 2. Health derivational morphology rules.

Derivational rule	**Example**
ose$(verb)→ osis$(noun)	sclerose(verb) → sclerosis(noun)
physeal$(adj) → physis$(noun)	adenohypophyseal(adj) → adenohypophysis(noun)
sis$(noun)→ ze$(verb)	dialysis(noun) → dialyze(verb)
a$(noun)→ iasis$(noun)	candida(noun) → candidiasis(noun)

Character level Grammar. We have identified a particular use of parentheses, square brackets and commas in the health domain. Examples of the use of parentheses and square brackets might be the following:

1. Sleep terrors [night terrors]
2. No Diagnosis or Condition on Axis I / No Diagnosis on Axis II [DSM-IV]
3. Premature (early) ejaculation
4. Trichotillomania (hair-pulling disorder)
5. Obstructive sleep apnea (adult) (pediatric)

In case 1, the square brackets are used to specify an equivalent expression of "*sleep terrors*". Similarly, in case 3 parentheses are used to indicate a synonym of "*premature*". Case 2 is different as brackets are used to point out the DSM version in which the description was included. In case 4 the content within parentheses categorises the kind of disorder that "*trichotillomania*" is. Finally, case 5 uses parentheses to indicate the domain to which the disorder is applicable, in that case to *adults* and *children*.

Similarly as in the previous cases, commas are utilised with different purposes in the medical knowledge. Below there are some examples:

1. Tobacco use disorder, Mild
2. Adverse effect of unspecified antidepressants, sequela
3. Circadian rhythm sleep disorder, shift work

In example 1, the comma is used to specify the degree of the disorder, whereas in example 2, it is used to define the kind of adverse effect. Finally, in example 3, the comma is used to specify the cause of the disorder.

This diverse use of parentheses, square brackets and commas, complicates labels, penalising matchers' performance. Thus, we decided to suppress commas and all content within parentheses and square brackets to avoid this penalisation. This simplifies labels and reduces irrelevant content. Nonetheless, in the future, we should investigate less aggressive solutions to reduce matchers penalisation while taking advantage of the content within parentheses.

5 Evaluation

The hypothesis has been evaluated by an experiment in which matchers with different configurations had to match several descriptions of the two most important classifications of diseases for mental health: the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) and the ICD-10. To evaluate the quality of the matchers, we used as gold standard the correspondences between both classifications published in DSM-5, where it is specified to which code in ICD-10 corresponds each description in DSM-5.

The input schemas were a source dataset with 200 entries randomly selected from DSM-5, and a target dataset with 177 descriptions included in ICD-10, which are the correspondences of the entries chosen from DSM-5.

The matchers selected were S-Match [14] and LogMap [15]. The main reasons of choosing these two matchers are their differences to carry out the matching process, and the diverse BK they use. Whereas the former carries out semantic matching, the latter is a highly scalable system that has reasoning and diagnosis capabilities allowing it to detect and repair unsatisfiability on the fly [10]. S-Match uses by default WordNet as BK, so it only includes general knowledge, whereas LogMap only incorporates by default biomedical knowledge provided by resources within of the Unified Medical Language System (UMLS). Regarding grammar, both matchers are limited to address general grammar. While S-Match includes tokenisation, lemmatisation and the translation of punctuation marks into logical connectives, LogMap implements string similarity measures, stop words elimination and word stemming.

The experiments were executed 4 times with each matcher, computing the standard metrics within the information retrieval community: *precision, recall* and *f-measure*. Firstly, with the vanilla version, which was our baseline in each case; secondly, with the lexicon extension; thirdly, with the grammar extension, and finally, with both extensions.

Figure 1 and figure 2 depict the results of the experiments executed in S-Match and LogMap, respectively. We can see how both matchers, S-Match and LogMap, improve their performance in terms of f-measure around 20% and 7% respectively. It is also noticeable, that overall both matchers achieve low results which are caused by the nature of the input labels, which on average are descriptions with more than 5-6 words, so this results in complex label formulas and low string similarity values.

Regarding S-Match, the vanilla version only has a general BK and the matcher is penalised mainly for the way in which it manages commas (each comma is considered as a disjunctive operator). This caused a huge number of false positives, which negatively affected precision, but also discovered, as side effect, a high number of correspondences, resulting in the highest recall. An example is the label "*Mild cognitive impairment, so stated*" which is transformed into the following label formula:

$$mild \ \& \ cognitive \ state \ \& \ impairment \mid state$$

From this label formula S-Match computes the following node formula:

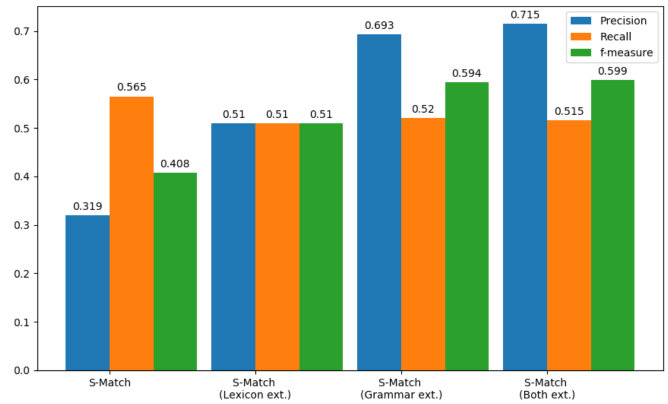

Fig. 1. Results of the experiments executed in S-Match.

$$(mild \mid state)\ \&\ (cognitive\ state \mid state)\ \&\ (impairment \mid state)$$

That means that if "*state*" has a relationship with a lemma within any label of the other ontology, the matcher will output a mapping even if the rest of the label is not related.

The lexicon extension considerably improves the performance (11%) by adding health lexicon knowledge, but this extension also avoids some of the correspondences discovered as side effect, mainly with the inclusion of lexical entries that were considered as single tokens in the vanilla version and now are compound tokens, so the recall slightly decreases.

The grammar extension is the one that drastically reduces the number of false positives mainly with the techniques applied at *phrase* and *character grammar levels* that were employed as a preprocessing step prior to the matching process. In addition, it also discovers new mappings thank to the derivational morphology implemented at *word grammar level*.

The combination of both lexicon and grammar extensions is the configuration that performs better in terms of f-measure, complementing each other and improving the baseline around 20%. However, the false positives of both extensions are also aggregated, being precision slightly penalised.

As for LogMap (see Figure 2), the vanilla version includes biomedical knowledge by default, resulting in a baseline with a performance over 60%.

The lexicon extension added knowledge coming from SPECIALIST, MeSH and WordNet, but it was the latter which produced the major impact as it added domain-independent knowledge contained in the labels. This new knowledge also produced some false positives, but on average this configuration improved the baseline around 6.3%. An example of false positive is: "*Narcolepsy without cataplexy but with hypocretin deficiency*" $\not\equiv$ "*Narcolepsy with cataplexy*", while an example of new true positive is: "*Acute stress disorder*" $\equiv$ "*Acute stress reaction*".

The grammar extension had a similar effect mainly because it also incorporated WordNet. In this case, tasks for *word* and *character grammar levels*

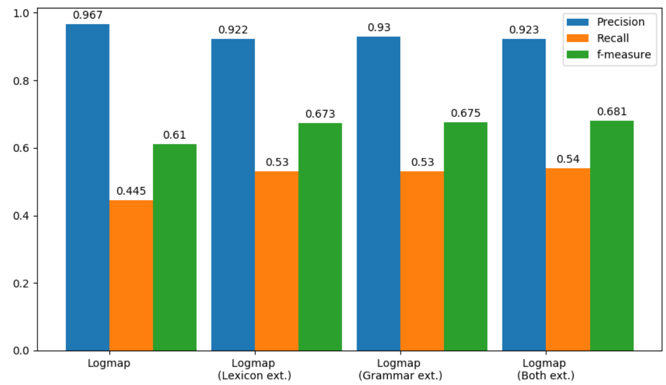

Fig. 2. Results of the experiments executed in LogMap.

had a low impact on LogMap. Nonetheless, *phrase level grammar* preprocessing had a significant impact, and the performance improved 6.5% with respect to the baseline. Examples of new true positives are: "*Trichotillomania (hair-pulling disorder)*" ≡ "*Trichotillomania*", and "*Overweight or obesity*" ≡ "*Obesity, unspecified*".

The combination of both extensions was the configuration that obtained the best performance, achieving the highest number of true positives discovered. In this case, the baseline is improved more than 7%.

6 Concluding Remarks

In this paper, we have presented an approach in which matchers can take advantage of both, domain lexicon and grammar to improve their performance when matching domain-knowledge resources. After evaluating our approach by matching some descriptions of mental health disorders included in DSM-5 and ICD-10 with S-Match and LogMap, we can conclude that our hypothesis is true, as both matchers improve their f-measure compared with the vanilla version.

It is interesting to highlight how the use of domain lexicon and grammar affects differently depending on the matcher. Whereas the domain lexicon extension has the major impact on LogMap, S-Match experiences its major improvement with the grammar extension. The main reason is that LogMap now can discover new mappings thank to domain-independent knowledge, and S-Match has label formulas significantly simplified. This information is useful in order to optimise efforts in the future, and help to decide whether is more valuable investing time focusing on integrating domain lexicon or grammar knowledge into matcher's KB.

As future work we should explore other factors that may affect matchers when matching domain-knowledge, such as the impact of each kind of knowledge represented within knowledge resources according to their levels of specificity. Moreover, it is interesting to delve into methods to aggregate lexicon and grammar results in order to optimise matcher's performance.

Acknowledgements

This research was partially supported by the European Commission with the grant agreement No. 607062 (ESSENCE Marie Curie ITN, http://www.essence-network.com/).

References

1. Amaro, R., Mendes, S.: Towards merging common and technical lexicon wordnets. In: Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon. pp. 147–160 (2012)
2. Annane, A., Bellahsene, Z., Azouaou, F., Jonquet, C.: Yam-bio–results for oaei 2017. In: CEUR Workshop Proceedings (2017)
3. Bella, G., Giunchiglia, F., McNeill, F.: Language and domain aware lightweight ontology matching. Journal of Web Semantics **43**, 1–17 (2017)
4. Bella, G., McNeill, F., Leoni, D., Quesada Real, F.J., Giunchiglia, F.: Diversicon: Pluggable lexical domain knowledge. Journal on Data Semantics pp. 1–16 (2019)
5. Bella, G., Zamboni, A., Giunchiglia, F.: Domain-based sense disambiguation in multilingual structured data. In: The Diversity Workshop at the European Conference on Artificial Intelligence (ECAI) (2016)
6. Bella, G., Elliot, L., Das, S., Pavis, S., Turra, E., Robertson, D., Giunchiglia, F.: Cross-border medical research using multi-layered and distributed knowledge (2020)
7. Browne, A.C., McCray, A.T., Srinivasan, S.: The specialist lexicon. National Library of Medicine Technical Reports pp. 18–21 (2000)
8. Da Silva, J., Revoredo, K., Baião, F., Euzenat, J.: A lin: improving interactive ontology matching by interactively revising mapping suggestions. The Knowledge Engineering Review **35** (2020)
9. Dragisic, Z., Ivanova, V., Li, H., Lambrix, P.: Experiences from the anatomy track in the ontology alignment evaluation initiative. Journal of biomedical semantics **8**(1), 56 (2017)
10. Euzenat, J., Shvaiko, P.: Ontology Matching - Second Edition. Springer (2013)
11. Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F., Couto, F.M.: The agreementmakerlight ontology matching system. In: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". pp. 527–541. Springer (2013)
12. Fellbaum, C.: WordNet. Wiley Online Library (1998)
13. Francopoulo, G.: Lmf iso 24613:2008 (Mar 2008)
14. Giunchiglia, F., Shvaiko, P., Yatskevich, M.: S-match: an algorithm and an implementation of semantic matching. In: European semantic web symposium. pp. 61–75. Springer (2004)
15. Jiménez-Ruiz, E., Cuenca Grau, B.: Logmap: Logic-based and scalable ontology matching. In: International Semantic Web Conference. pp. 273–288. Springer (2011)
16. Lipscomb, C.E.: Medical subject headings (mesh). Bulletin of the Medical Library Association **88**(3), 265 (2000)
17. Quesada Real, F.J., McNeill, F., Bella, G., Bundy, A.: Improving dynamic information exchange in emergency response scenarios. In: Proceedings of 14th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2017). pp. 824–833 (2017)
18. Quesada Real, F.J., McNeill, F., Bella, G., Bundy, A.: Identifying semantic domains in emergency scenarios. In: 15th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2018). pp. 1130–1132 (2018)