Aligning the Bulgarian BTB WordNet with the Bulgarian Wikipedia

Kiril Simov, Petya Osenova, Laska Laskova, Ivajlo Radev, Zara Kancheva
IICT-BAS, Sofia, Bulgaria
{kivs|petya|laska|radev|zara}@bultreebank.org

Abstract

The paper reports on an ongoing work that manually maps the Bulgarian WordNet BTB-WN with Bulgarian Wikipedia. The preparatory work of extracting the Wikipedia articles and provisionally relating them to the WordNet lemmas was done automatically. The manual work includes checking of the corresponding senses in both resources as well as the missing ones. The main cases of mapping are considered. The first experiments of mapping about 1000 synsets show the establishement of more than 78 % of exact correspondences and nearly 15 % of new synsets.

1 Introduction

There is still lack of sufficient knowledge for solving many important NLP tasks, such as word sense disambiguation (WSD), relation extraction, named entity linking, event detection, etc. Up to now a number of attempts have been provided in the community that integrate various linguistic and semantic resources in smart ways. These are, among others, SemLink (Palmer, 2009), Predicate Matrix (de Lacalle et al., 2014), UBY (Gurevych et al., 2012), BabelNet (Navigli and Ponzetto, 2012). SemLink combines PropBank (Kingsbury and Palmer, 2002), VerbNet (Kipper-Schuler, 2005), and FrameNet (Baker, 2008). Predicate Matrix extends SemLink with a mapping from its lexical units to WordNet synsets (Fellbaum, 1998). UBY was created for two languages — English and German. It combines WordNet and GermaNet with Wiktionary, Wikipedia, FrameNet and VerbNet for English, and Wiktionary and Wikipedia for German. BabelNet also combines many multilingual resources including WordNet and Wikipedia. All these examples demonstrate two facts: (1) a single knowledge resource is not sufficient for the most of the NLP tasks; and (2) the automatic integration of the various distinct resources is error prone. This is especially true for low-resource languages that totally miss such resources or their existing resources are rather small in size.

Here we report on an effort to integrate Bulgarian WordNet (BTB-WN) (Osenova and Simov, 2018) with the Bulgarian Wikipedia. We are considering mapping of two semantic objects — *concepts* (meaning expressed by common words) and instances of such concepts called *named entities*. The integration is meant to be performed manually in order to ensure high quality of the result. The integrated knowledge graph will include the current version of BTB-WN extended with: a) new senses and new synonyms for the existing synsets — all extracted from the articles in the Bulgarian Wikipedia; b) a controlled number of named entities that are specific to Bulgaria and c) increasing the number of terminological concepts in various domains. Thus the integrated resource will combine general lexica with encyclopedic knowledge (terminology).

The expected result would be twofold: a) the mutual enrichment and improvement of both resources and b) handling of WSD in a more effective way by integrating the encyclopedic knowledge from Wikipedia and the lexical information from WordNet.

The structure of the paper is as follows: in the next section related work is presented. Section 3 outlines the approach to the mapping as well as the results. The last section concludes the paper.

2 Related Work

Needless to say, one of the most notable resources that link WordNet and Wikipedia is BabelNet — an automatically created very large, wide-coverage multilingual semantic network (Navigli and Ponzetto, 2012). BabelNet encodes knowl-

edge as a labeled directed graph. It is created by linking the largest multilingual Web encyclopedia – Wikipedia, to the most popular computational lexicon — WordNet. BabelNet has been built in 3 steps. The first step was to automatically combine WordNet and Wikipedia by mapping the WordNet senses to Wikipedia articles. The second step was to collect the multilingual lexicalizations of the BabelNet synsets by using human-generated translations. These translations were provided by Wikipedia as well as by a machine translation system for translating the occurrences of the concepts within sense-tagged corpora. The third step was to establish relations between the Babel synsets through collecting all the relations found in WordNet together with all Wikipedias in the languages of interest. The integration was performed by an automatic mapping and by filling the lexical gaps in resource-poor languages with the aid of Machine Translation. The result is an "encyclopedic dictionary" that provides concepts and named entities lexicalized in many languages and connected with large amounts of semantic relations.

In spite of having at disposal such a resource as BabelNet, our motivation to invest efforts in mapping the WordNet to the Wikipedia was as follows: a) adding more locally important content into the existing mappings and b) enriching the resource that was constructed automatically with validated data. The Babelfy service is very good at detecting concepts and names (given the availability of relevant data per language), but it still has problems with disambiguation among local people or places with the same name, or between a concept and a name. For example, the verb литва (litva, "start to fly") is identified only as the country Литва (Litva, "Lithuania") whose graphical form coincides with the verb; similar for the adjective русия (rusiya, "blond") and the name of the country Русия (Rusiya, "Russia").

In (Osenova and Simov, 2018) Osenova and Simov mention the initial attempt for annotating of named entities (NE) in Bulgarian Treebank (BulTreeBank) with URIs from DBpedia. This process was done with the same goal, namely to extend BTB-WN in two directions: (1) the number of senses for lemmas that are already in BTB-WN; and (2) the instances of the concepts. However, the BulTreeBank appeared to contain only a small number of named entities. Thus the extension was insufficient and it required the use of the Wikipedia URIs and DBpedia classes for the missing NEs. The authors also report on the automatic extension of BTB-WN with automatically derived Bulgarian synsets on the basis of the English ones through the usage of the English Wiktionary. After manual checking of around 11000 suggestions, BTB-WN was enriched by around 5000 synsets.

(McCrae, 2018) reports on the manual mapping of the Princeton WordNet (PWN) instances to the English Wikipedia. He proposes that a subset of PWN instance concept synsets is automatically linked and manually evaluated on Wikipedia articles in order to "provide a gold standard for link discovery". This is done by matching PWN lemmas to all Wikipedia titles containing the lemma. Then by using a special tool, human annotators evaluate the links. This tool shows the PWN definitions and the first paragraph from the Wikipedia article so the annotators are able to confirm or reject the mapping. The same paper also suggest 5 types of links between PWN and Wikipedia: exact — one synset to one article; broad — several synsets to one article; narrow — one synset to several articles; related — one-to-one relation, but not the same concept; unmapped — not possible to map. This method proved to be highly successful and even yielded a report with 8 errors which aimed to improve PWN. We follow very closely the approach of this work except that we are interested in mapping not only the instances, but all possible lemmas in BTB-WN.

(Rudnicka et al., 2017) present another attempt at linking two large lexico-semantic databases, namely the Princeton WordNet of English and the plWordnet of Polish language. The approach considers models and ideas originating from the bilingual lexicography and translation studies. For the creation of the PlWordnet language data from contexts of use attested in large language corpora was used rather than from dictionaries and the approach focused on word uses, not concepts.

A synset in PWN is viewed as a representation of a lexicalised concept, while in plWordNet it is a set of lexical units sharing constitutive lexico-semantic relations and features. The synset includes such lexical units that share a set of lexico-semantic relations, called constitutive relations (hyper/hyponymy, holo/meronymy, type/instance, etc.). In some cases the constitutive relations might be irrelevant, so constitutive features are also used – stylistic register, aspect,

semantic classes of verbs and semantic classes of adjectives. Glosses, examples and substitution tests are also applied in the plWordnet. The mapping strategy refers to the synset level and includes looking for pairs of plWordnet and PWN synsets that are close in meaning. The stages of the mapping are as follows: an analysis of the sense and relation structure of a source synset, the selection of candidates for a target synset, the choice of a target synset and an inter-lingual relation that links the source and target synsets. Having in mind the complex schema of mapping between the two WordNets we doubt that such a mapping could be successfully established between the WordNet and the Wikipedia even for the one and the same language. Expectedly, when named entities are highly predominant in the mapping, we might envisage also a high number of exact mappings, but for common words this is not so straightforward. For that reason, we decided to perform the mapping manually. For the first step our goal was to extend BTB-WN with new synsets, synonyms and mappings to Bulgarian Wikipedia.

Another approach that could be taken into account when aiming to extend the WordNet is its alignment with a FrameNet (if such a resource has been constructed for the language). A recent and rather innovative example of the development of a FrameNet based on a corpus of written Dutch, and annotated with PropBank predicates and roles is the project of (Vossen et al., 2018). In this project the creation of the FrameNet also exploits already manually classified data about real world events which specify frame constraints on the described situations. This data is manually related to texts describing the events. In future work we will consider this approach to extend the coverage of BTB-WN as well as to add new constraints on the combinations of the senses within texts.

3 BTB-WN to Wikipedia Mapping

In this section we present the correspondences between the synsets within BTB-WN and the pages from the Bulgarian Wikipedia.

3.1 Wikipedia Page to Synset Correspondence

The first step was to establish a correspondence between lexical entries in BTB-WN and the Bulgarian Wikipedia. For each lemma within BTB-WN we automatically selected all the articles in Wikipedia that match that lemma. In order to do this, the article titles were cleaned from the modifiers given in brackets like in the following example: the lemma маса (masa) corresponding to "table" (a piece of furniture), "mass" (a body of matter), and "mob" (a disorderly crowd of people) is mapped to Wikipedia articles with titles like: Маса "Mass" (physical term); Маса (мебел) "Table (furniture)", etc. The special disambiguation articles play an important role like Маса (пояснение) in this example. Their importance comes from the fact that they provide additional information about the potential synonyms. Such an example in this case is the connection from Маса to Заземяване "Ground (electricity)" which was a missing sense within the current version of BTB-WN. For each Bulgarian Wikipedia article we also extracted the title of the corresponding English article in order to facilitate the process of selecting the right meaning and the process of mapping between BTB-WN and the English WordNet.

After the extraction of the relevant Wikipedia pages we grouped together the pages corresponding to a given lemma and all the BTB-WN synsets that contain the lemma. These groups have been represented in XML and loaded into CLaRK System[1] for inspection and mapping. A screen shot of the data loaded in the system is presented in Fig. 1. Each group is represented via `<eq>` element. In the representation we use the tree layout settings of the system in order to present not only the structure elements but also their content. Each group contains one or more pages, thus one or more entries for the same lemma. If an entry contains more than one lemma, this entry will be added to several groups if there are appropriate Wikipedia pages. In the figure we can observe two expanded groups — one for the lemma "Iceberg" and one for the lemma "Aquarium". For each page the layout shows the Bulgarian title of the page, then the English title (if there is a link to an English Wikipedia page). Thus, the annotator[2] could understand the sense described by the Wikipedia article without expanding the structure of the page. Of course, if necessary, the annotator could read more from the content of the page. For each entry

[1] For a description see (Simov et al., 2004b). The system could be downloaded from `http://bultreebank.org/en/clark/`.

[2] We call the people that manually establish the mapping between the two resources *annotators*, but a more appropriate term is necessary such as *mappers* or *knowledge relaters*.

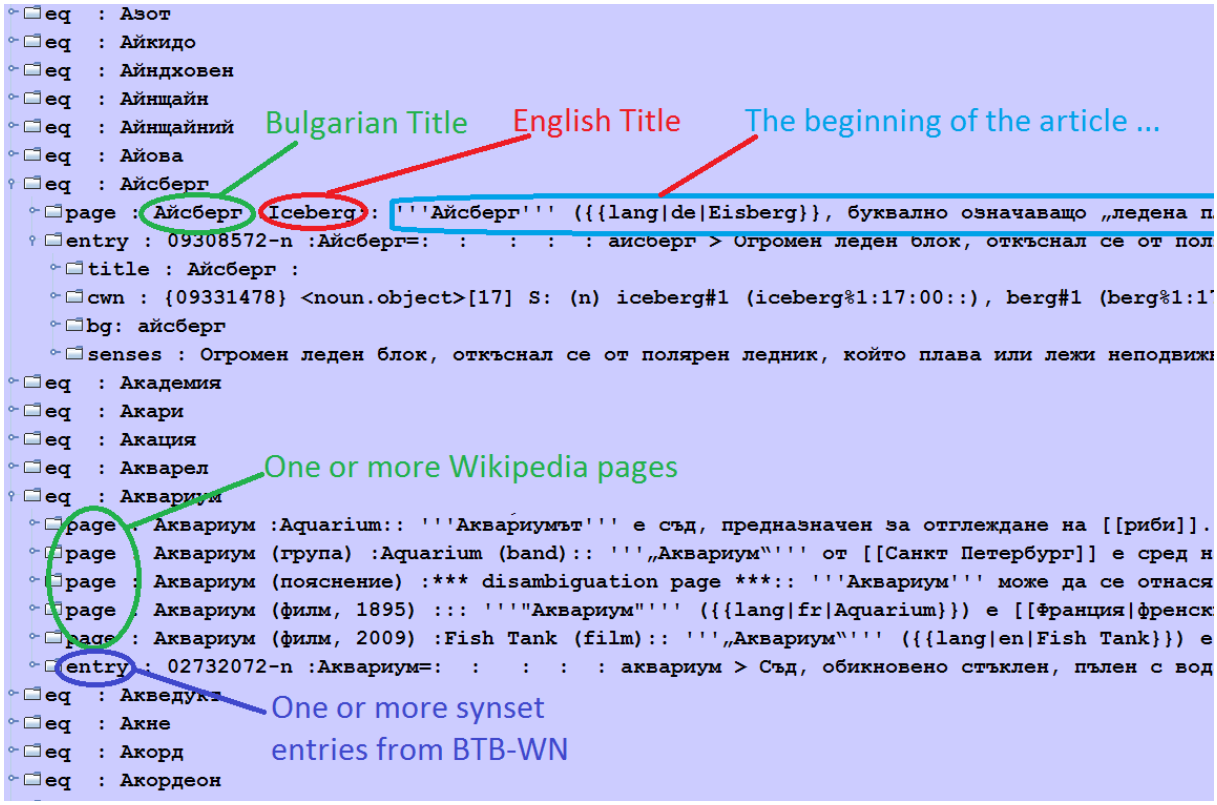

Figure 1: Representation of the groups of matched Wikipedia pages and BTB-WN synsets (represented via <entry> element.)

the layout shows the PWN identifier; the mapping to Wikipedia page (if such has been selected); the list of lemmas for the synset; and finally the definition related to the synset. Again, the annotator might read the important information without expanding the structure of the entry. In the example of the group for "Iceberg" the structure of an entry is as follows. The element <cwn> contains the mapping information to PWN. The element <bg> contains the list of lemmas of the synset. The element <senses> contains one or more definitions (if selected from different sources) and zero or more examples of uses of the lemmas in the corresponding sense.

The group for "Iceberg" represents the simplest case of one-to-one mapping. The actual connection is established by copying the title of the appropriate Wikipedia page as a first element of the entry. The group for "Aquarium" demonstrates the case when more than one Wikipedia page corresponds to a given lemma. Here we have a page corresponding exactly to an entry in BTB-WN. Several pages exist for named entities like a band, two movies – one French and one British. Also there is a disambiguation page, marked with "*** disambiguation page ***". In cases of disambiguation page we also added the pages that are mentioned within the disambiguation list. Similarly, we add the redirect pages pointing to some of the other pages within the group. In some cases such redirect pages provide synonyms or derivative lemmas. In this way we try to provide as much information as possible from the Bulgarian Wikipedia to the annotator.

Following the mapping strategy, mentioned above, for about 22 000 synsets in BTB-WN we extracted a little more than 13 000 Wikipedia articles. For each sense (sense in BTB-WN is defined as a lemma in some of the synsets) in BTB-WN the annotators received a list of the corresponding Wikipedia articles. Thus they were able to check whether the selected sense is presented within Wikipedia and to establish correspondence if it is the case. After consulting the individual senses in BTB-WN, the annotators checked whether new meanings had to be added to it. The new meaning could be a sense for the common word or a named entity. In both cases the annotator created a new lexical entry in BTB-WN.

3.2 Named Entities Processing

Because of the high productivity in the case of named entities, many common words are presented as named entities in Wikipedia. Since our main goal was to introduce more locally centered names, these respectively were considered as important. Thus, the annotator first filtered the candidates in order to introduce only the important names. More specifically, we defined names of importance in the following way:

- As a first step, only names of persons, organizations and locations are considered;
- For location names we select names of Bulgarian places or of well-known foreign places;
- For the rest of the names only well-known names are considered.

Although this definition is not very precise, it helped us to filter quite a lot of location named entities. Here we additionally introduced a restriction to include larger cities in Europe (larger than 100 000 citizens if they are not well-known). In this way for example, Шенген ("Schengen") is included in BTB-WN although it has less than 4000 citizens, but Буден ("Boden", a city in Sweden) is not included although its transliteration in Bulgarian coincides with an adjective. In our future work we need to make the definition more precise in order to cover all the names in Wikipedia, but without overloading the WordNet with the ambiguity coming from very rare named entities.

The above selection criteria are to some extent arbitrary[3]. For example, for some countries the limit of 100 000 citizens is too restrictive. Especially for small countries or countries in Europe. For other countries this might allow many not well known cities. In order to provide an additional evaluation of the importance of the named entities, we use a gazetteer created during the development of the BulTreeBank Pipeline for Bulgarian — see (Simov et al., 2004a) and (Savkov et al., 2012) and during the compilation of the Bulgarian treebank (2001-2004). The names in it were collected from the following sources: (1) Bulgarian law documents containing the names of all villages, towns, cities, municipalities in Bulgaria; (2) Names from touristic advertisements; and (3) list of names manually selected from a ranking list of potential named entities from a large corpus of Bulgarian. We consider the names in the gazetteer as representative for Bulgarian texts. They also contains all Bulgarian location names. The gazetteer contains more than 26 000 records, but some of them are not basic forms (lemmas) because during the preparation of the gazetteer we selected non-basic forms like vocatives, plurals and definite forms.

All the Wikipedia pages were extracted that correspond to the names in the gazetteer. We extracted 10 899 pages altogether. From them 1 515 pages were already extracted on the basis of the lemmas within BTB-WN. Thus we marked there 1 515 as important, but still the annotators could select names that are marked in this way. The rest 9 384 pages were classified as Bulgarian locations, other locations, people, organizations and other. They will be checked for inclusion in BTB-WN at a later stage. In this way we selected also some important names that are not considered at the beginning of this work.

3.3 Mapping Cases

Here we consider different cases of correspondence among pages and entries, grouped together on the basis of the lemmas from BTB-WN. Each annotator was instructed to check the aligned WordNet synsets and the Wikipedia articles for the following cases:

- Exact mapping of senses represented in both resources;
- A concept represented in Wikipedia, but not in WordNet. In such a case they had to create a new synset and to establish a mapping;
- An admissible named entity in Wikipedia, missing in WordNet. In such a case they had to create a new synset and to establish a mapping.

Whenever a new synset was created, it was also mapped to the corresponding PWN synset when possible (for more details see (Osenova and Simov, 2018)). The annotation was performed by 5 people that considered nearly 1000 WordNet lemmas, automatically mapped to more than 1300 Wikipedia articles. Table 1 presents the distribution of the different cases.

The first category (first line — None) contains the number of no correspondences between the

[3] As it was pointed to us by one of the reviewers.

Correspondence	Number Total: 1309	%
None	276	21.08
Equality	688	52.57
Many to One	128	09.78
New Concept	128	09.78
New Named Entity	68	05.19
New Synonyms	21	01.60

Table 1: Percentage of the different cases.

two resources. In this case none of the Wikipedia articles describes a synset in BTB-WN. The reason for this usually is the named-entity-centered nature of Wikipedia. For example, under the title Плейбой "Playboy" Wikipedia has only one article on the Playboy journal. In BTB-WN there is an entry corresponding to PWN synset with a gloss "a man devoted to the pursuit of pleasure." The closest page in English Wikipedia is "Playboy lifestyle" which requires a more complex mapping. Such a page is missing in the Bulgarian Wikipedia. Similarly, the word Стожер (stozher) in the Wikipedia is only a name of a village and a newspaper, while WordNet records only the concept стожер (stozher) as pillar. Thus, the WordNet entity cannot be mapped to Wikipedia. This case corresponds to McCrae's *Unmapped links*.

The second category (Equality) describes the equality relation, where both resources describe the same concept. For example, Столица (stolitsa), "capital" is defined in the same way in both resources. These cases are the majority of all mappings. It corresponds to McCrae's *Exact links*.

The third category (Many-to-One) presents the case where different parts of the same Wikipedia article are dedicated to different concepts. Often, but not always, this is the case for the disambiguation pages. Among the concepts, one usually corresponds to the mapped WordNet synset. For example, in Wikipedia, Стойка (Stoyka) has several representations as a given name or a surname, but it also refers to the concept of (body) posture and the concept of stand. BTB-WN contains only one concept — that of the posture. Another problem in this case is that the two pages for these general concepts do not exist, but they are defined only in the disambiguation page. Thus, the annotator has to use a special relation to the disambiguation page. The Индекс (indeks), "index" page illustrates another example that is treated in a similar manner. In this case, the authors of the Wikipedia page point out that the word индекс might refer to several things, among which a list of items, a superscript or subscript character, a hierarchical classifier, and a value on a measurement scale. Two of these concepts are lexicalized as индекс in the WordNet and they are mapped to the article with a Many-to-One relation, which corresponds to McCrae's *Broad links*.

In both cases, the annotator has to perform one more operation before moving on, that is, to check whether the BTB-WN does not already contain the seemingly missing concepts; it might be the case that they are lexicalized in a different way, i.e. in other terms. Here, the annotators rely on information from Wikipedia, and, of course, on their own linguistic competence. Whenever deemed necessary, and especially when dealing with terminological units, they consult a synonym dictionary or a thesaurus. Needless to say, there would be two possibilities: a) the right match is found, or b) not found, because it is missing. In the Стойка example, the concept for "stand" was already present in the WordNet, so the annotator established a Many-to-One correspondence between the article and the synset, and added the term стойка to the set of synonyms. In the Индекс example, the new concepts found in Wikipedia were indeed missing from the BTB-WN and thus the annotator created two new synsets mapping them to the article with a Many-to-One relation.

In some cases, the new concept introduced in the Wikipedia article, is given only a short definition and the term is linked to an empty page. Given the dynamic nature of Wikipedia, we decided to map this type of pages to the corresponding BTB-WN synsets with an additional *empty* relation; from here we can expect one of the two positive outcomes — on the one hand, the annotators are free to contribute to the Bulgarian Wikipedia by providing new content (a time-consuming task which at this point is given a low priority), an on the other hand, we keep the possibility of future resource enrichment by not excluding a potentially useful mapping.

The fourth and the fifth categories (New Concept and New Named Entity) correspond to the case in which the Wikipedia article introduces one or more new concepts — both types or instances. We can distinguish several cases here.

The Wikipedia article lists some or all of the hy-

ponyms of the concept named in the title. For example, Абак (abak), "abacus", contains information about the different types of abacuses. Each of these types prompts the creation of a new synset. In this cases we reuse the definitions from the Wikipedia article. We also select examples from the article. This allows for BTB-WN to be used independently from Wikipedia.

The Wikipedia article is dedicated to different concepts which are not linked by a hypernym — hyponymy relation. This type of relatedness corresponds to McCrea's understanding of *Related links*. The nature of the relatedness remains unspecified but the new concept is always linked to some existing one in the WordNet: through homonymy, derivation, systematic polysemy, semantic expansion, etc. Let us give some examples.

- The article Авария (avariya) describes a technogenic disaster. It is related to the synset авария, катастрофа (breakdown, equipment failure) by a causal link.

- The article Инвалидност (invalidnost), disability is related to the synset инвалид (invalid), "disabled person" derivationally. Here we annotate the mapping as derivational, but in future we will add more specific relations depending on the semantic relation.

- As for the systematic polysemy, two are the most common types.
  The first one regards the relation between a title understood as "an identifying appellation signifying status or function", and the person who is given this title because they have the corresponding status or function. As a rule, the Wikipedia article describes the title while the existing WordNet concept is related to the person. The annotators create a new synset linked to the page with an Equality relation and also indicate the specific type of relatedness between the preexisting synset and the page.
  The second type of systematic polysemy is characteristic of some geographical named entities, such as Бахамски острови (Bahamski ostrovi, Bahamas). This multiword expression has two meanings. It can refer to the country, the Commonwealth of the Bahamas, or to a geographical region, in this case the island group known as Lucayan Archipelago. The annotators apply the same strategy as the one described above.

The Wikipedia article introduces a hypernym. For example, Камион (kamion), "truck" in Wikipedia is a hyperonym of the two synsets for truck and van, presented within the current version of BTB-WN.

The sixth category (New Synonyms), features the case when the corresponding synset is part of the WordNet, but there are some missing synonyms that come from the Wikipedia. For example, the multiword expression Кралство Камбоджа, "Kingdom of Cambodia" is missing in the synset that contains the name of Cambodia.

As it can be seen, in more than 78 % of the cases we establish a correspondence between synsets in BTB-WN and the Bulgarian Wikipedia. In our view this is a good coverage. Also we have added about 15 % new concepts and named entities.

4 Conclusion

The paper presents our initial attempts in enriching BTB-WN with mappings to the Bulgarian Wikipedia. The first annotation results are promising in showing that WordNet profits well from this mapping — especially in adding synonyms, new senses and new instances.

The importance of such a resource is envisaged at least in the following directions: enhancing named entity linking, relation extraction and word sense disambiguation of high quality for tasks, involving Bulgarian data. The mapping also provides access to the whole Wikipedia articles which could contribute valuable information for the usage of the corresponding concepts and named entities.

The main source of enriching BTB-WN appeared to be the named entities and the domain terms. We also noticed that Wikipedia is a valuable resource for including MWEs — predominantly terminological units, but not only. Since the named entities are too many, as mentioned above, we focused on local ones because they are important for processing Bulgarian data, and also — they can be viewed as a valuable localized supplementary contribution to BabelNet.

Another issue is that Wikipedia contains mainly nouns. Thus, the mappings enriched the noun network and the instances of names. For the verbs, adjectives and adverbs other enriching sources

should be considered. Through the derivation relations in WordNet, however, we still could incorporate the presented in Wikipedia deverbal and adjectival nouns.

In future work we envisage to map BTB-WN also to other semantic resources such as Wikidata. We have started with Wikipedia because it provides more human oriented information which facilitates the mapping. In addition, Wikidata is heavily extracted from Wikipedia and we hope this to allow for an easy mapping.

In the long run, we envisage also incorporating more Bulgarian concepts and named entities with the idea to construct a Bulgarian knowledge graph aligned to linguistic knowledge — senses and grammatical features.

Acknowledgements

This research was funded by the Bulgarian National Science Fund grant number 02/12/2016 — *Deep Models of Semantic Knowledge (DemoSem)*. The contribution of Ivajlo Radev and Zara Kancheva has been partially supported by the Bulgarian Ministry of Education and Science under the National Research Programme "Young scientists and postdoctoral students" approved by DCM # 577 / 17.08.2018. We are grateful to the anonymous reviewers for their valuable remarks, comments, and suggestions. All errors remain our own responsibility.

References

Collin Baker. 2008. FrameNet, present and future. In Jonathan Webster, Nancy Ide, and Alex Chengyu Fang, editors, *The First International Conference on Global Interoperability for Language Resources*, Hong Kong. City University, City University.

Maddalen Lopez de Lacalle, Egoitz Laparra, and German Rigau. 2014. Predicate matrix: extending semlink through wordnet mappings. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 903–909, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Christiane Fellbaum, editor. 1998. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA ; London, May.

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 2012. Uby - a large-scale unified lexical-semantic resource based on lmf. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 580–590, Avignon, France, April. Association for Computational Linguistics.

Paul Kingsbury and Martha Palmer. 2002. From treebank to propbank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands - Spain, May. European Language Resources Association (ELRA).

Karin Kipper-Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

John P. McCrae. 2018. Mapping WordNet Instances to Wikipedia. In *In: Proceedings of Ninth Global WordNet Conference*, pages 62–69. The Global WordNet Association.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and applicationof a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Petya Osenova and Kiril Simov. 2018. The Data-driven Bulgarian WordNet: BTBWN. *Cognitive Studies | Études cognitives*, 18(1713).

Martha Palmer. 2009. Semlink: Linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*, page 9–15.

Ewa Katarzyna Rudnicka, Maciej Tomasz Piasecki, Tadeusz Piotrowski, Łukasz Grabowski, and Francis Bond. 2017. Mapping WordNets from the perspective of inter-lingual equivalence. *Cognitive Studies | Études cognitives*, 17(1373):1–17.

Aleksandar Savkov, Laska Laskova, Stanislava Kancheva, Petya Osenova, and Kiril Simov. 2012. Linguistic Processing Pipeline for Bulgarian. In *Proceedings of LREC 2012*, pages 2959–2964.

Kiril Simov, Petya Osenova, Sia Kolkovska, Elisaveta Balabanova, and Dimitar Doikoff. 2004a. A Language Resources Infrastructure for Bulgarian. In *Proceedings of LREC 2004*, pages 1685–1688.

Kiril Simov, Alexander Simov, Hristo Ganev, Krasimira Ivanova, and Ilko Grigorov. 2004b. The CLaRK System: XML-based Corpora Development System for Rapid Prototyping. *Proceedings of LREC 2004*, pages 235–238.

Piek Vossen, Antske Fokkens, Isa Maks, and Chantal Van Son. 2018. Open Dutch Framenet. In Tiago Timponi Torrent, Lars Borin, and Collin F. Baker, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France, may. European Language Resources Association (ELRA).