Mapping WordNet Instances to Wikipedia

John P. McCrae

Insight Centre for Data Analytics
National University of Ireland Galway
Galway, Ireland
`john@mccr.ae`

Abstract

Lexical resource differ from encyclopaedic resources and represent two distinct types of resource covering general language and named entities respectively. However, many lexical resources, including Princeton WordNet, contain many proper nouns, referring to named entities in the world yet it is not possible or desirable for a lexical resource to cover all named entities that may reasonably occur in a text. In this paper, we propose that instead of including synsets for instance concepts PWN should instead provide links to Wikipedia articles describing the concept. In order to enable this we have created a gold-quality mapping between all of the 7,742 instances in PWN and Wikipedia (where such a mapping is possible). As such, this resource aims to provide a gold standard for link discovery, while also allowing PWN to distinguish itself from other resources such as DBpedia or BabelNet. Moreover, this linking connects PWN to the Linguistic Linked Open Data cloud, thus creating a richer, more usable resource for natural language processing.

1 Introduction

Princeton WordNet (Fellbaum, 2010; Miller, 1995, PWN) and Wikipedia, especially in machine readable form such as DBpedia (Lehmann et al., 2015), are two of the most widely used resources in natural language processing. The nature of these resources is distinct, with WordNet constituting a lexicon of words in the English language and Wikipedia being an encyclopedia describing entities in the world. This means that WordNet should contain all the common nouns, verbs, adjectives and adverbs and Wikipedia should contain the proper nouns referring to notable entities in a text. However, in fact there is a significant overlap between these two resources as Wikipedia contains pages for abstract general concepts, such as "play"[1], while PWN contains many proper nouns for concepts such as *Paris*, for which PWN has four synsets for the city in France (`i83645`), the city in Texas (`i84698`), the mythological prince (`i86545`) and a plant (`i102495`). In the case of WordNet, the choice of which proper nouns to include has had certain biases, for example there are many synsets for cities in the United States, e.g, *Paterson, New Jersey* (`i84527`), but not for *Kawasaki*, a city in Japan that is ten times larger. If however, PWN were to expand to include more proper nouns, it would lead to a much larger resource that would overlap significantly in its coverage with DBpedia. In fact, there have been several attempts to automatically create such a resource, most notably BabelNet (Navigli and Ponzetto, 2012) and UBY (Gurevych et al., 2012), however these resources have to rely on automatic alignment of the concepts. Instead, we propose that the concepts for named entities can be mapped to Wikipedia and that these concepts can thus be removed or replaced with links in future versions of PWN. Since PWN is created by careful manual effort, it is clear that an automatic mapping would not be compatible with the nature of PWN. Instead, as a principal contribution of this paper, we present the first manually created mapping between PWN instances and Wikipedia articles. This could be further used to link PWN to other resources including WikiData and GeoNames as well as help in the automatic translation of parts of WordNet.

In this paper, we first define the scope of the problem, in particular in terms of the number of instances and proper nouns that exist in PWN and

---

[1] `https://en.wikipedia.org/wiki/Play_(activity)`

their distribution. We then review some existing work on mapping PWN and Wikipedia. We present our method of linking, that uses Wikipedia categories to propose an alignment between sets of concepts simultaneously and the tool we created based on this that allows our annotators to quickly map the concepts between one resource and another. Finally, we present the results of our annotation, in particular in terms of the total effort and work required to create this mapping and conclude with some discussion and analysis of the results.

2 On Proper Nouns in WordNet

Princeton WordNet is a lexicon, that consists of a graph of *synsets*, which are collections of words that are synonymous, linked by a number of properties. All words in a synset have the same part-of-speech, however unfortunately there is only a single category for nouns and in fact synsets may contain a mixture of proper and common nouns, e.g., *Caterpillar,cat* (`i51642`). The links in the graph are of different types and the link `instance_hypernym` links a synset to a concept that is an *instance of* (Miller and Hristea, 2006), giving a limited set of proper nouns that we can systematically identify. There are in total 7,742 synsets in PWN which are instance hypernyms of 946 synsets and these will be the main focus of our work. Of these nearly all contain words starting with a capital letter, and of the 16 that don't, can be explained as follows: 7 are not capitalized for orthographic reasons, e.g., *al-Muhajiroun*, 6 should be capitalized but are not in WordNet, e.g., *pampas*, 2 should not be instance hypernyms but instead normal hypernyms *isle,islet* (`i85598`) and *sierra* (`i86184`) and 1 *church mouse* (`i48540`) is likely erroneous. As such, we can say that the set of synsets that are marked as instance hypernyms of a concept are all named entities in the world. However, there are many other synsets that contain one or more capitalized word as an entry and it is clear that we are not capturing all the proper nouns in PWN. In particular, there are a large number of capitalized words that refer to names of species or other terms in the Linnaean Taxonomy, e.g., *Felis catus* or *genus Hydrangea* and these are not instances of another synset and often share a synset with common nouns, e.g., *domestic cat,house cat,Felis domestics,Felis catus* (`i46594`). In addition, there are several other large categories of proper

nouns that are not captured by this approach especially beliefs, e.g., *Buddhism* (`i79765`) and languages, e.g., *German,High German,German language* (`i73125`). However, simply using the capitalization to detect proper nouns produces a lot of false positives, including acronyms and terms including a proper noun such as *Scotch terrier, Scottish terrier, Scottie* (`i46443`). As such, for this work we have focussed only on the synsets which are instances of synsets, as these are the terms that seem to be most encyclopedic in their content. A breakdown of the major synsets is given in Figure 1, and as we can see the major categories are (`i35562`), which is named people, (`i35580`), which is named places. A few other categories that have large number of entities include rivers and other geological features ((`i85104`),(`i85439`) and (`i85674`)), gods (`i86570`), events, especially wars (`i35586`), social groups, such as terrorist organizations (`i79103`) and books (`i69848`).

3 Related Work

The goal of mapping WordNet to Wikipedia has been recognized as an important one, however most of the focus has so far been on the automatic creation of mappings between the two resources, and this has led to the creation of wide-coverage lexicons that are useful for NLP applications but cannot act as a gold standard for NLP in the same way that WordNet does. The most notable such resources is BabelNet (Navigli and Ponzetto, 2012), whose mapping of WordNet to Wikipedia is based on the use of a word-sense disambiguation algorithm, where contexts are created for the Wikipedia and WordNet entities by means of using the surrounding synsets and the article texts. A second step then selects the highest scoring mapping based on structuring the Wikipedia page content using WordNet relations. The authors report a maximum F-Measure of 82.7% with a precision of 81.2%, showing that while BabelNet is a high-quality resource, it cannot be considered a gold standard. This method improved on a previous approach by these authors (Ponzetto and Navigli, 2009), which used the taxonomic structure of the resources. Another method to link WordNet and Wikipedia has been through Personalized Page Rank (Agirre and Soroa, 2009), which was first attempted as a method for linking these resources in (Toral et al., 2009) and then was fur-

```
i35545 - entity (7742)
│
├─ i35546 - physical entity (6616)
│   │
│   ├─ i35548 - thing (415)
│   │   │
│   │   └─ i85104 - water;body of water (411)
│   │
│   └─ i35549 - object;physical object (6199)
│       │
│       ├─ i35550 - unit;whole (3666)
│       │   │
│       │   └─ i35552 - animate thing;living thing (3339)
│       │       │
│       │       └─ i35553 - organism;being (3339)
│       │           │
│       │           └─ i35562 - soul;someone;individual;somebody;mortal;person (3320)
│       │
│       ├─ i35580 - location (2109)
│       │
│       ├─ i85439 - formation;geological formation (143)
│       │
│       └─ i85674 - earth;land;dry land;terra firma;ground;solid ground (279)
│
└─ i35547 - abstract entity;abstraction (1126)
    │
    ├─ i35574 - psychological feature (720)
    │   │
    │   ├─ ...
    │   │   │
    │   │   └─ i86570 - divinity;deity;god;immortal (311)
    │   │
    │   └─ i35586 - event (223)
    │
    ├─ i35577 - attribute (30)
    │   │
    │   └─ i35582 - time (28)
    │
    ├─ i35589 - group;grouping (154)
    │   │
    │   └─ i79103 - social group (142)
    │
    ├─ i35593 - communication (192)
    │   │
    │   └─ i69848 - written communication;black and white;written language (159)
    │
    └─ i35594 - amount;quantity;measure (29)
        │
        └─ i108052 - fundamental quantity;fundamental measure (28)
```
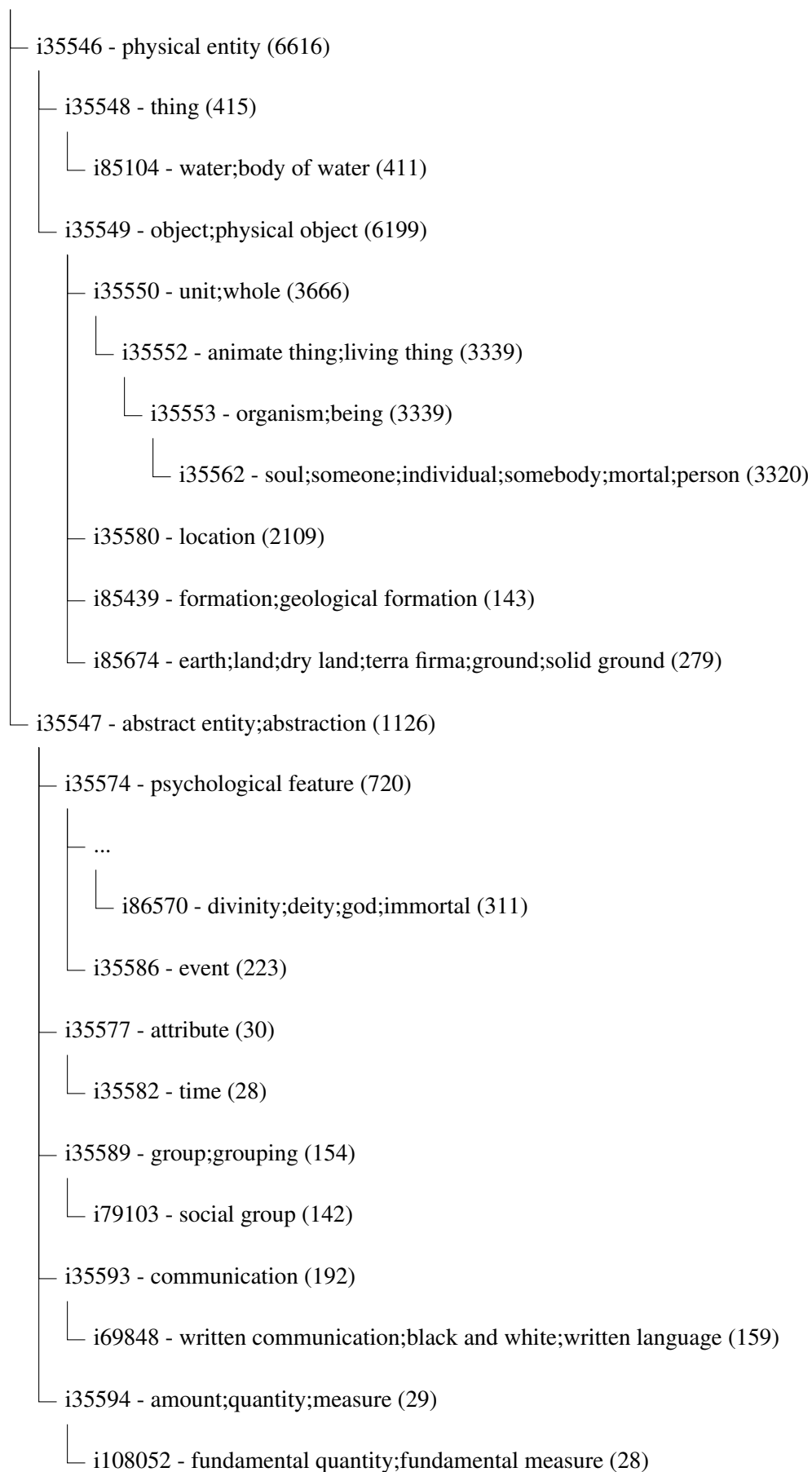
Figure 1: The most frequent hypernyms of instances in Princeton WordNet

ther improved by (Niemann and Gurevych, 2011), by the introduction of "thresholds". Niemann and Gurevych's methodology forms the basis of the UBY resource (Gurevych et al., 2012). Finally, Fernando and Stevenson (Fernando and Stevenson, 2012) proposed using semantic textual similarity methods and showed results that obtained an F-Measure of 84.1% outperforming Ponzetto and Navigli's approach. Notably, this work also created a gold standard of Wikipedia-WordNet mappings that can be used for evaluation of further approaches to linking. However, this mapping is only of 200 words and as such is not on the same scale as the resource introduced in this paper.

Another large-scale resource that has been constructed by combining WordNet and Wikipedia is Yago (Suchanek et al., 2008; Suchanek et al., 2007), which created an ontology of concepts created from Wikipedia categories. This showed a very high accuracy in the mapping of concepts (97.7%), however this does not deal with the actual entities as in this work.

WordNet has also been linked to a number of other lexical resource by a variety approaches, including SemCor (Mihalcea and Moldovan, 2000), where texts were annotated with WordNet synset identifiers and this was used as a basis to create links to other resources including FrameNet (Baker et al., 1998, FN) and VerbNet (Schuler, 2005, VN), which were linked in (Shi and Mihalcea, 2005). Another linking was created by the SemLink (Palmer, 2009; Bonial et al., 2013), also based on the annotation of a corpus with PWN, FN and VN. Finally, mappings have also been proposed between WordNet and Wiktionary[2], a free dictionary from the WikiMedia Foundation, in works such as (McCrae et al., 2012) and (Meyer and Gurevych, 2011).

4 Mapping WordNet to Wikipedia

Our goal is to create a large manual mapping between a subset of Princeton WordNet and Wikipedia, however simply identifying this subset and starting annotation is not a suitable approach as looking up each WordNet synset in Wikipedia and recording the results would be a slow and dull process. We could try to improve this by matching the lemmas of WordNet entries to the titles of Wikipedia articles, but this would have a very low coverage as the article title for a Wikipedia

article must be unique so often includes specific disambiguating terms. To expand the coverage of this we consider a WordNet lemma to match a Wikipedia article if it matches the title ignoring case before the first comma or parentheses or any page that redirects to this article. Thus, we would match the lemma "Paris" to the page titles "Paris", "Paris, Texas" and "Paris (Mythology)". In addition, we also included information from disambiguation pages, as collected by DBpedia (Lehmann et al., 2015)[3]. This method captures most of the mappings as only 77 WordNet synsets have no candidates in Wikipedia, however it also creates significant ambiguity with an average of 21.6 candidates for each synset. For these reasons, we try to resolve these differences by suggesting category mappings, inspired by (Suchanek et al., 2008).

4.1 Unambiguous Category Matches

We start by considering all pairs of WordNet instance synsets and Wikipedia articles as $W = \{s_i, a_j\}$. Let all hypernyms of a synsets be the set of $H(s_i)$ and let all categories for a Wikipedia article by $C'(a_j)$. We also consider all categories of categories and all categories of those categories to create a list of categories $C(a_j)$, as the categories for some articles can be very narrow. The set of mappings between non-instance synsets and Wikipedia categories is created as follows:

$$M = \{h, c | \exists \{s_i, a_j\} \in W : h \in H(s_i) \land c \in C(a_j)\}$$

This creates a very large number of mappings and we wish to choose which mappings are most suitable, thus we create a score to rank them. We use two main constraints to do this, firstly, we note that short lemma matches tend to be quite ambiguous, e.g., "Paris, Texas" is less ambiguous than "Paris", and secondly, we notice that mappings that create a lot of duplicate matches are challenging to annotate. Firstly, we define $l(s_i, a_j)$ as the follows, where $L(s_i, a_j)$ is the set of matching terms between the WordNet instance and the Wikipedia article, $t(l)$ gives the length (number of tokens) of this matching terms in this mapping and α is a constant:

$$l(s_i, a_j) = \sum_{l \in L(s_i, a_j)} t(l) - \alpha.$$

Secondly, we generate a set of proposed mappings based on a hypernym, $h \in H(s_i)$ and a Wikipedia category $c \in C(a_j)$ as follows

$$P(h,c) = \{(s,a) | h \in H(s) \wedge c \in C(a) \\ \wedge L(s,a) \neq \emptyset\}$$

We say that a pair (s,a) is *unambiguous* in $P(h,c)$ if there is no distinct element $(s',a') \in P(h,c)$ such that $s = s'$ or $a = a'$. Finally, we score a mapping as follows:

$$s(h,c) = \sum_{(s,a)\in P(h,c)} \sigma(s,a)$$

$$\sigma(s,a) = \begin{cases} l(s,a) & \text{if } (s,a) \text{ is unambiguous} \\ & \text{in } P(h,c) \\ -\beta & \text{otherwise} \end{cases}$$

For parameters we chose $\alpha = 1$, as this allows us to ignore mappings created from single tokens and $\beta = 10$ as this provided a good trade-off between allowing some ambiguity in the mappings. In fact, the first 2,500 entries were annotated with a higher β value, but it become clear that this was too strict so we permitted more ambiguity in the mapping.

4.2 Annotation Tool

In order to create the annotations a tool was created to show the proposed mappings, which is depicted in Figure 2. This tool shows all the proposed category mappings and then all the individual instances and Wikipedia articles that will be linked. For each WordNet instance the definition in WordNet is given and for the Wikipedia article, its first paragraph is given. For each case, we selected whether the mapping was valid and then submitted the proposed mapping. The system allows two extra actions, "Reject", which is the same as unselecting all mappings and submitting the form and "Reject Wikipedia Category", which removes all mappings involving this Wikipedia category. This option was introduced as some Wikipedia categories were clearly not likely to map to any synsets in Wikipedia [4].

---

[4] An example is https://en.wikipedia.org/wiki/Category:Timelines_of_cities_in_France

5 Resource and Evaluation

We used the above described methodology to annotate the vast majority of the mappings (7,582 mappings), while the remaining 239 synsets had no good candidates in Wikipedia, principally due to spelling variants and this includes the 77 synsers with no candidates and other synsets for which the category approach did not work. These remaining 239 synsets were then mapped directly (on a spreadsheet). We also used this pass to sort the links into the following types:

Exact The WordNet synset and Wikipedia article exactly describe the same entity.

Broad The Wikipedia article describes several things, of which the entity described by the WordNet synset is only one of. An example of this is the Wikipedia article for the "Wright Brothers"[5], which is linked broader to two WordNet synsets for each brother. In this case, Wikipedia redirects "Orville Wright" and "Wilbur Wright" to this article.

Narrow The opposite of 'broad', i.e., the WordNet synset describes multiple Wikipedia articles. An example is *Rameses, Ramesses, Ramses* (i96663) defined as "any of 12 kings of ancient Egypt between 1315 and 1090 BC"[6], while each is a separate Wikipedia article.

Related The Wikipedia article does not describe the WordNet synset but something intrinsically linked to it, and the lemmas of the WordNet synset have redirects to this article. For example *Hoover, William Hoover, William Henry Hoover* (i95579) is mapped to "The Hoover Company" describing the company he founded. Wikipedia also redirect "William Hoover" to this article.

Unmapped A small number of entities in WordNet were not possible to map to Wikipedia, either because the synset was not in Wikipedia (this was the case for many terrorist organizations), the description and name did not match anything in Wikipedia (for a

---

[5] https://en.wikipedia.org/wiki/Wright_brothers

[6] This is also an error as there were only 11 Egyptian pharoahs named Ramesses

Proposed Mapping

WordNet
wn31-00001740-n entity

that which is perceived or known or inferred to have its own distinct existence (living or nonliving)

Wikipedia
People by country of descent

Reject

Reject Wikipedia Category

WordNet	Wikipedia	Accept
(wn31-10953409-n) Princess of Wales, Lady Diana Frances Spencer, Diana, Princess Diana *English aristocrat who was the first wife of Prince Charles; her death in an automobile accident in Paris produced intense national mourning (1961-1997)*	Diana, Princess of Wales *Diana, Princess of Wales (Diana Frances; née Spencer; 1 July 1961 – 31 August 1997), was the first wife of Charles, Prince of Wales, who is the eldest child and heir apparent of Queen Elizabeth II. Diana was born into a family of British nobility with royal ancestry as The Honourable Diana Spencer. She was the fourth child and third daughter of John Spencer, Viscount Althorp and the Honourable Frances Roche. She grew up in Park House, situated on the Sandringham estate, and was educated in England and Switzerland. In 1975, after her father inherited the title of Earl Spencer, she became Lady Diana Spencer.*	☑
(wn31-10953680-n) Duchesse de Valentinois, Diane de Poitiers *French noblewoman who was the mistress of Henry*	Princess Charlotte, Duchess of Valentinois *Princess Charlotte of Monaco, Duchess of Valentinois (Charlotte Louise Juliette Grimaldi de Monaco; 30 September 1898 – 15 November*	☑

Figure 2: The Annotation Tool used to create the mappings

Exact	Broad	Narrow	Related	Unmapped
7,582	54	21	30	59

Table 1: The size of the resource by type of link

few place names) or the synset was not something that would generally be in Wikipedia, e.g., different names for gods, such as *Jupiter Fidius, Protector of Boundaries* (i86982)

We used the following heuristic to help with this mapping. If the Wikipedia page title exactly matched one of the lemmas or the Wikipedia article was of the form "X, Y" or "X (Y)" and X was one of the lemmas and Y occurred in the definition of the synset, we accepted it as an exact match[7]. For example, this allowed us to easily validate the mappings for the Wikipedia articles "Paris" (the capital of France), "Paris, Texas" and "Paris (mythology)". All other mappings (1,733) were manually assigned one of the above cate-

gories. As a result of this mapping process we also detected 56 errors (0.7%) and improved 11 mappings, by which we mean that we changed a broader/narrower link to an exact link. For example, the synset *Downing Street* (i83390), was moved from "10 Downing Street" to "Downing Street". The complete size of each of these categories is given in Table 1, in a few cases a wordnet synset was mapped using "narrower" to multiple Wikipedia articles thus the 7,742 entities created 7,746 links.

5.1 Improvements to Princeton WordNet

In the process of creating the mappings between PWN and Wikipedia, we closely studied a section of Princeton WordNet and thus found a large number of errors within the resource. As such we submitted a report to the developers of Princeton WordNet detailing the following errors[8]:

- Two synsets were identified to be duplicates (referring to the same concept).

- One synset was suggested to be split

---

[7]As an aside, this heuristic of matching the diffentiating part of the title to the WordNet definition may have been quite effective for establishing mappings in Section 4.1, but was not considered until most of the mapping was completed. In this paper, we focus on the construction of the resource and describe the methodology we followed.

[8]This document may be viewed at `https://docs.google.com/document/d/1yn-UurCoeuKk_OwRzDaj1dYW88k2l0ymD7YtBIiVlCM`

- 17 lemmas with typos were detected

- Two links were found to be incorrect

- Four synsets described concepts for which no reference could be found outside of PWN

- 41 definitions were found to be factually inaccurate, this was mostly due to the year that a person was born in or died in not being correct.

- We suggest 1,062 new synset members to be added to existing synsets. These were derived from the Wikipedia page titles and so represent standard well-attested variants of existing names. These primarily consist of variations of names, e.g., "University of Cambridge" is the official name for *Cambridge,Cambridge University* (i51397), but in some cases are more significant, e.g., *Seward's Folly* (i41225) is more commonly known as the "Alaska Purchase".

5.2 Resource

The mapping has been created and is made available from the following URL[9]. In addition, the mapping will be contributed to the Global WordNet Index (Bond et al., 2016; Vossen et al., 2016) and as a mapping to the DBpedia project[10]. In this case, we provide an RDF file that links the Global WordNet ILI URIs with DBpedia URIs. The mapping is made available under a CC-Zero license to enable its re-use in as many places as possible. The source code for tools used in this project are available on GitHub [11].

6 Conclusion

We have presented a new mapping of all the instances in WordNet to Wikipedia articles. This represents the largest gold standard mapping for tasks such as link discovery (Nentwig et al., 2017) and is likely to be a basic resource for many tasks in natural language processing. For the future development of Princeton WordNet as a resource, this mapping can form the basis by which PWN can distinguish itself from an encyclopedia, by replacing the instance links with direct links to Wikipedia. Moreover, by linking to Wikipedia articles, we can further link to many other resources, for example it is only a matter of changing the URL to find a DBpedia entity that can be used to find machine readable information about the data. Furthermore, all Wikipedia articles are now linked to WikiData entities, so we can easily find that *Paris, City of Light, French capital, capital of France* (i83645) is linked to WikiData entity Q90[12] and then this can give us identities in many other databases including GeoNames (2968815), OpenStreetMap (71525) and even the official Twitter account (@Paris). Finally, it is worth noting that Wikipedia and Wikidata also contains links to these concepts in other languages, and as such, this linking can create a partial translation of a section of WordNet. As such, this transforms WordNet into a richer linked resource that can be part of the Web of Linguistic Linked Open Data (McCrae et al., 2016).

Acknowledgments

References

[Agirre and Soroa2009] Eneko Agirre and Aitor Soroa. 2009. Personalizing pagerank for word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 33–41. Association for Computational Linguistics.

[Baker et al.1998] Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.

[Bond et al.2016] Francis Bond, Piek Vossen, John P. McCrae, and Christiane Fellbaum. 2016. CILI: the Collaborative Interlingual Index. In *Proceedings of the Global WordNet Conference 2016*.

[Bonial et al.2013] Claire Bonial, Kevin Stowe, and Martha Palmer. 2013. Renewing and revising semlink. In *The GenLex Workshop on Linked Data in Linguistics*.

[Fellbaum2010] Christiane Fellbaum. 2010. WordNet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

---

[9] http://jmccrae.github.io/wn-wiki-instances/ili-map-dbpedia.ttl
[10] http://github.com/dbpedia/links
[11] https://github.com/jmccrae/wn-wiki-instances

[12] https://www.wikidata.org/wiki/Q90

[Fernando and Stevenson2012] Samuel Fernando and Mark Stevenson. 2012. Mapping WordNet synsets to Wikipedia articles. In *Proceedings of The Eighth International Conference on Language Resources and Evaluation*, pages 590–596.

[Gurevych et al.2012] 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. Association for Computational Linguistics.

[Lehmann et al.2015] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. 2015. DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2):167–195.

[McCrae et al.2012] John McCrae, Elena Montiel-Ponsoda, and Philipp Cimiano. 2012. Integrating WordNet and Wiktionary with lemon. *Linked Data in Linguistics*, pages 25–34.

[McCrae et al.2016] John P. McCrae, Christian Chiarcos, Francis Bond, Philipp Cimiano, Thierry Declerck, Gerard de Melo, Jorge Gracia, Sebastian Hellmann, Bettina Klimek, Steven Moran, Petya Osenova, Antonio Pareja-Lora, and Jonathan Pool. 2016. The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud. In *10th Language Resource and Evaluation Conference (LREC)*.

[Meyer and Gurevych2011] Christian M Meyer and Iryna Gurevych. 2011. What psycholinguists know about chemistry: Aligning Wiktionary and WordNet for increased domain coverage. In *IJCNLP*, pages 883–892.

[Mihalcea and Moldovan2000] Rada Mihalcea and Dan Moldovan. 2000. Semantic indexing using WordNet senses. In *Proceedings of the ACL-2000 workshop on Recent Advances in Natural Language Processing and Information Retrieval*, pages 35–45. Association for Computational Linguistics.

[Miller and Hristea2006] George Miller and Florentina Hristea. 2006. WordNet nouns: Classes and instances. *Computational Linguistics*, 32(1):1–3.

[Miller1995] George A Miller. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

[Navigli and Ponzetto2012] Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

[Nentwig et al.2017] Markus Nentwig, Michael Hartung, Axel-Cyrille Ngonga Ngomo, and Erhard Rahm. 2017. A survey of current link discovery frameworks. *Semantic Web*, 8(3):419–436.

[Niemann and Gurevych2011] Elisabeth Niemann and Iryna Gurevych. 2011. The people's web meets linguistic knowledge: automatic sense alignment of Wikipedia and Wordnet. In *Proceedings of the Ninth International Conference on Computational Semantics*, pages 205–214. Association for Computational Linguistics.

[Palmer2009] Martha Palmer. 2009. Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the generative lexicon conference*, pages 9–15. Pisa Italy.

[Ponzetto and Navigli2009] Simone Paolo Ponzetto and Roberto Navigli. 2009. Large-scale taxonomy mapping for restructuring and integrating Wikipedia. In *IJCAI*, volume 9, pages 2083–2088.

[Schuler2005] Karin Kipper Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

[Shi and Mihalcea2005] Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. *Computational linguistics and intelligent text processing*, pages 100–111.

[Suchanek et al.2007] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706. ACM.

[Suchanek et al.2008] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2008. YAGO: A large ontology from Wikipedia and WordNet. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(3):203–217.

[Toral et al.2009] Antonio Toral, Oscar Ferrández, Eneko Agirre, Rafael Munoz, Informatika Fakultatea, and Basque Country Donostia. 2009. A study on linking Wikipedia categories to WordNet synsets using text similarity. In *Proceedings of the 2009 Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 449–454.

[Vossen et al.2016] Piek Vossen, Francis Bond, and John P. McCrae. 2016. Toward a truly multilingual Global Wordnet Grid. In *Proceedings of the Global WordNet Conference 2016*.