Inclusion of Lithological terms (rocks and minerals) in The Open Wordnet for English

Alexandre Tessarollo, Alexandre Rademaker
Petrobras, IBM Research and FGV/EMAp
alexandretessarollo@gmail.com, alexrad@br.ibm.com

Abstract

We extend the Open WordNet for English (OWN-EN) with rock-related and other lithological terms using the authoritative source of GBA's Thesaurus. Our aim is to improve WordNet to better function within Oil & Gas domain, particularly geoscience texts. We use a three step approach: a proof of concept-level extension of WordNet, a major extension on which we evaluate the impact with positive results and a full extension encompassing all GBA's lithological terms. We also build a mapping to GBA which also links to several other resources: WikiData, British Geological Survey, Inspire, GeoSciML and DBpedia.

Keywords: wordnet, rocks, lithology, domain extension, geology, NLP

1. Introduction

Oil & Gas Exploration and Production companies annually invest billions of dollars gathering documents such as reports, scientific articles, business intelligence articles and so on. These documents are the main base for major decisions such as whether to drill exploratory wells, bid or buy, production schedules and risk assessments (Rademaker, 2018). However, most of the processing of this fundamental data is still done by human professionals actually reading it rather than by a computational system. Considering that this unstructured data is growing exponentially, management of such data and finding relevant content quickly has become one of companies and professionals most critical challenges (Antoniak et al., 201 6; Schoen et al., 2018). Even though Natural Language Processing (NLP) has significantly advanced over the past years, the specific domain of Oil & Gas has its own challenges, some of them presented in (Rademaker, 2018).

Assessing geosciences papers one can notice that among the most common properties raised are usually geographic location (Palkowsky, 2005), geological time and lithological information. In a previous work (Rademaker et al., 2019) we addressed some of the issues regarding geological time. In this work we approach the lithological information aspect.

Section 2. gives a brief description of similar projects. Section 3. present our authoritative source for terms and definitions. Section 4. shows our platform of choice for extending the WordNet. In section 5. we present and discuss the proposed changes. In section 6. we raise some relevant and recurrent issues we faced and the reasoning supporting our decisions. Section 7. presents some comparative statistics over a given corpus processed both with the original WordNet and our extended version. Section 8. sums up the results and points to future works.

2. Related works

Princeton WordNet (PWN) (Fellbaum, 1998a) does not cover many terms and concepts specific to certain domains as pointed out by (Buitelaar and Sacaleanu, 2002), hence the need to expand PWN for each domain in order to tap into its potential as a NLP resource (Amaro and Mendes, 2012). WordNet extensions for specific domains are relatively common.

Medical WordNet (MWN) (Smith and Fellbaum, 2004) reviews PWN medical terms through a corpus which includes a validated corpus of sentences involving specific medically relevant vocabulary. The corpus is composed by the definitions of medical terms already existing in WordNet, sentences generated via the semantic relations in PWN and sentences derived from online medical information services targeted to consumers. BioWN (Poprat et al., 2008) was another attempt to extend WN to the biomedical domain from the Open Biomedical Ontologies (OBO). OBO would provide terms, definitions and relations to be included in WN. According to the authors, the attempt failed due to issues on several softwares and resources that eventually prevented the success of the initiative. (Buitelaar and Sacaleanu, 2002) leans on German's compositional aspect to extend GermaNET with medical terms. The relevance of the candidate terms is then measured in a given domain corpora. Roughly the definitions arise from the compositional rule used to build the term in the first place.

In the legal domain, JurWN (Sagri et al., 2004) builds upon the Italian ItalWordNet (IWN) database, aiming to extend it to the legal domain. IWN (Roventini et al., 2003) is the Italian component of the EuroWordNet (Vossen, 2002). Words were selected from frequent terms used in queries of the major legal information retrieval systems, while definitions were taken from handbooks, dictionaries, legal encyclopedias and other main technical concepts. The LOIS (Lexical Ontologies for legal Information Sharing) project (Peters et al., 2006) encompass legal WordNets for six different languages (Italian, Dutch, Portuguese, German, Czech, English) based on the EuroWordNet framework. It used a subset of JurWN as a seed and added new terms on the basis of authoritative resources, national and EU legislative text and legal text.

GeoNames WordNet (GNWN) (Bond and Bond, 2019) links the GeoNames[1] geographical database to wordnets in different languages. GeoNames provides both the terms and definitions to be included in GNWN as an instance of a given synset (e.g.: Paris as an instance of city).

Noticeable from all these initiatives is the approach consid-

[1] https://www.GeoNames.org/

Proceedings of the Workshop on Multimodal Wordnets (MMWN-2020), pages 33–38
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020
© European Language Resources Association (ELRA), licensed under CC-BY-NC

ered to extend a wordnet to a given domain. Some refer to a corpus (custom built or pre-existing material) to gather a list of words to include in the wordnet, and then to an authoritative material such as dictionaries and encyclopedias for the definitions. Others refer to authoritative material that have both terms and definitions, such as ontologies.

3. INSPIRE and GBA's Thesaurus

The Infrastructure for Spatial Information in the European Community (INSPIRE) (Parliament and of the Council, 2007) was created to build upon existing resources (infrastructure and data) of the Member States. The original focus is to support EU policies and activities which may have an impact on the environment. Particularly within the scope of this work, Inspire offers an organized codelist for lithology[2]. This resource is actually maintained by the Geological Survey of Austria (Geologische Bundesanstalt) within its "GBA Thesaurus" (GBA). Regarding lithology, GBA presents a richer material than Inspire, all accessible online[3] and available for download[4].

GBA is an ontology based on the Simple Knowledge Organization System (SKOS) vocabulary (Isaac and Summers, 2009). Each term has a Universal Resource Identifier (URI) and is related to other terms via SKOS object properties. Within the scope of our work, we have *broader* and its counterpart *narrower*. Therefore, "mammal has *broader* animal" and "animal has *narrower* mammal". GBA follows SKOS convention to only assert direct hierarchical links. The name of the term is given by *prefLabel* data property, while the definition is given by *definition* data property. String values are given in English as well as in German. GBA uses a few other SKOS properties like *related match*, *close match*, *hidden label* and others. Particularly *exact match* is used to map GBA to other resources, INSPIRE included. The downloadable material for GBA is a Resource Description Framework[5] (RDF) file, which means it is organized in triples consisting of subject, predicate and object.

At its description, GBA states that Lithology comprises loose- and bed-rock, classified according to their modal composition and grain size, respectively. Magmatic-, polygenetic-, metamorphic- and fault-rocks are classified based on International Union of Geological Sciences (IUGS) recommendations[6]. Sedimentary rocks classifications refer to international standards. Considering GBA alignment with IUGS recommendations and its mapping to WikiData[7], British Geological Survey (BGS)[8], Inspire[9], GeoSciML[10] and DBpedia[11], i.e. several governmental, multinational and community consensual based open-source initiatives, we assumed GBA's thesaurus for lithology as an authoritative figure. Therefore, it is not scope of this work to question the correctness of GBA's material, but to map it into the WordNet.

4. Princeton WordNet and the Open Wordnet for English

Princeton WordNet (PWN)[12] (Fellbaum, 1998b; Miller et al., 1990) is a large lexical database of English and one of the most widely-used language resources in natural language processing. It works well as a dictionary and a thesaurus for uses of English, as found, for instance, in newspapers and general knowledge texts, such as Wikipedia. Unfortunately, its development came to a halt over a decade ago.

In (Muniz et al., 2018) some of the authors present previous initiative to expand PWN with geological terms. This work started as fork of PWN release 3.0. Initially, PWN was converted to a human-readable text format and later an Emacs[13] mode and a validation tool were developed. It is called Open Wordnet for English (OWN-EN) and maintained at `http://github.com/own-en/`. The focus is on the expansions of PWN to specific domains (mainly geology and its intersection with Oil & Gas exploration) but also on the fixing of well-known bugs founded in PWN over the years. In this repository one can find the products of this paper, i.e., the extended WN as well as the mapping between it and GBA.

In the future, we aim to consider the merge of our OWN-EN with the Open English WordNet (McCrae et al., 2019). This is another fork of PWN being developed under an open source methodology. Its 2019 release fixed over 3,500 errors in PWN. The authors are committed to release new versions at least every year. One can contribute to the project and/or use its products at `https://en-word.net`.

5. Extending OWN-EN from GBA's Thesaurus

WordNet's cornerstone is its several types of conceptual relations. Of our interest, we have the *hyponym of* (counterpart *hypernym of*), which indicates a subtype relation. The *part holonym of* (counterpart *part meronym of*) indicates a component relation. Similarly, *substance holonym of* (counterpart *substance meronym of*) indicates a component relation for substances. The *Domain of synset - topic* (counterpart *domain of synset - member*) indicates the topic a given concept (synset), as in "geology is *domain of synset - topic* of rock".

From the GBA thesaurus, we consider the labels and definitions of the concepts and the concepts relations. But GBA's definitions were not taken literally since they were

[2] `http://inspire.ec.europa.eu/codelist/LithologyValue`

[3] `https://thesaurus.geolba.ac.at`

[4] `https://github.com/schmar00/gba-thesaurus/tree/master/rdf`

[5] `https://www.w3.org/TR/rdf-concepts/`

[6] `https://www.iugs.org/history`

[7] `https://www.wikidata.org/wiki/Wikidata:Main_Page`

[8] `http://data.bgs.ac.uk`

[9] `http://inspire.ec.europa.eu/codelist/LithologyValue`

[10] `http://resource.geosciml.org/classifier/cgi/lithology`

[11] `https://wiki.dbpedia.org`

[12] `https://wordnet.princeton.edu`

[13] `https://www.gnu.org/software/emacs/`

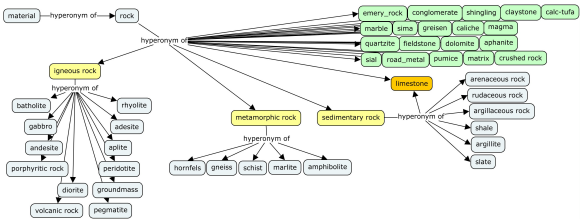

Figure 1: Rock in WordNet

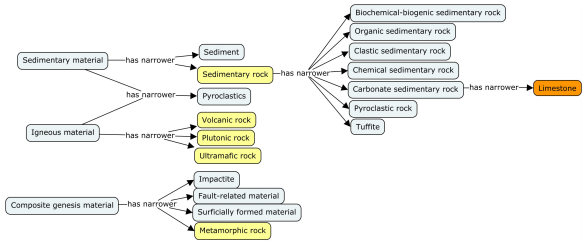

Figure 2: Rock in GBA

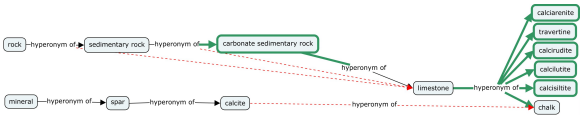

Figure 3: Limestone relations in WN: red ones to be removed, green ones to be included

not written as dictionary definitions. For instance, they include many in-depth descriptions of the concepts and references to scientific literature. Our goal was to provide for the new synsets, as much as possible, Aristotelian definitions following general lexicography methodology. Besides all information from GBA incorporated into our OWN-EN, we also provide a mapping from GBA concepts URIs to the OWN-EN sense keys. This will also facilitate future revisions of our resource once new releases of GBA are made available. Because GBA is already mapped to multiple other resources (WikiData, BGS, Inspire, GeoSciML and DBpedia), our mapping encompasses these resources as well.

In WN, the word *rock* has many senses, and the one that resembles the geological meaning is 14696793-n (rock : material consisting of the aggregate of minerals like those making up the Earth's crust). The reader should consider this sense wherever rock is mentioned henceforth. Figure 1 shows how *rock* is represented in WN, while figure 2 shows a few of the uppermost lithologies in GBA. A first look at both shows that WN has at least some hierarchical issues: there are nineteen synsets (in green) that are *hyponym of rock* instead of one of the three main WN's classes of rock: igneous, metamorphic and sedimentary (all in yellow). Finally, there is *limestone* (in orange): *hyponym of* both *rock* and *sedimentary rock*. Considering *sedimentary rock* is *hyponym of rock*, the *limestone* to *rock hyponym of* is at least redundant.

In yellow in figure 2 we can see that *sedimentary rock* and *metamorphic rock* are represented in both WN and GBA. WordNet's *igneous rock* has three counterparts in GBA: *volcanic rock*, *plutonic rock* and *ultramorfic rock*. Finally, *limestone* in GBA is *hyponym of carbonate sedimentary rock* which in turn is *hyponym of sedimentary rock*. Notice that GBA does not have a term for 'rock' pure and simple. Instead its top concepts are three types of material and from those arise different rocks and other materials. 'Rock' however is used to define other ones (see *sedimentary rock* below). Due to this and to the fact that *rock* is a relevant term in everyday language, we chose to keep this WN synset, add the three top concepts of GBA and allocate GBA's specific terms downwards from these four synsets.

To expand and adapt WN onto lithology domain we used GBA's terms and properties starting from the different types of rocks and lithologies. The obvious choice for mapping SKOS relationships to WN relationships is as first discussed in (van Assem et al., 2006). In our case, where in GBA A has *broader* B, in WN we defined A as *hyponym of* B; likewise, where in GBA B has *narrower* A, in WN we defined B as *hypernym of* A. For the sake of simplicity, we'll use WN's relations names henceforth. We also opted for lower case terms when changing or adding synset in WN.

GBA does not have explicit relations between rocks and the minerals that compose it, but we inferred the rock compositions in WN relations *substance holonym of* and *substance meronym of* from GBA's definitions. We also used WN's *domain of synset - TOPIC* and *member of this domain - TOPIC*, as explained later on.

As a proof of concept of our approach, we worked with *limestone* and initially analyzed only definitions and *hypernym of* and *hyponym of* relations. Afterwards we worked on the *substance holonym of*, *substance meronym of*, *domain of synset - TOPIC* and *member of this domain - TOPIC* relations. While the first step enriches WN with lithological terms, the second step ventures into the mineral domain, expanding WN even further. Once we set this work routine, we expanded the task to include all *carbonate sedimentary rock* and *clastic sedimentary rock*, the main types or reservoir rocks for Oil & Gas, ergo the most relevant for this industry. Finally, we included all of GBA lithology ontology into WN.

In WN *limestone* has the aforementioned redundant relations between *rock* and *limestone*. These and other deleted relations are highlighted in red in figure 3. In green the inclusion of 6 new terms and their 18 new relations with other terms. Note that due to the inclusion of *carbonate sedimentary rock* between *sedimentary rock* and *limestone* the *hypernym of* and *hyponym of* relations between *limestone* and *sedimentary rock* are no longer necessary.

For the six new terms added to WN we used the GBA definitions with minor adjustments in order to get closer to Aristotelian definitions and general lexicography methodology. For the ones that already existed in WN, a careful analysis was necessary and carried out top to bottom.

In GBA the concept *sedimentary rock* is defined as *a rock formed from post depositional consolidation of sediments (by processes of compaction, cementation, crystallization, or biogenic binding)* and it is a *hyponym of sedimentary*

material [14]. Analyzing both definitions and comparing with WN's definition for *sedimentary rock* [15] we conclude that, as explained in Section 6., WN's current definition for *sedimentary rock* is technically poor and should be replaced.
The other words WN already had were *limestone* and *chalk*. *Chalk* was classified as a mineral in WN, but GBA states that *chalk* is a rock and that rocks are composed of minerals. WN had 14806598-n (chalk : a soft whitish calcite), while GBA defines it as *a light-coloured (white-gray) marine limestone composed almost entirely of fine crystalline calcite. These porous limestones consist of foraminifera and calcareous algae, and usually contain chert nodules*. On this term we discarded WN's current definition and replaced it with GBA's.
As for *limestone* WN has 14936226-n (limestone : a sedimentary rock consisting mainly of calcium that was deposited by the remains of marine animals). The fragment *a sedimentary rock* is represented in the hypernyms of relations *limestone* $\rightarrow$ *carbonate sedimentary rock* $\rightarrow$ *sedimentary rock*; the fragment *consisting mainly of calcium* will be addressed by a meronym relation; finally, *that was deposited by the remains of marine animals* is not mentioned by GBA's definition. The first two parts can be removed without losses. As for the last part, (Encyclopaedia Britannica, 2018) states *limestone has two origins: (1) biogenic precipitation from seawater, the primary agents being lime-secreting organisms and foraminifera; and (2) mechanical transport and deposition of preexisting limestones, forming clastic deposits*. Therefore, the whole WN definition for limestone can be disregarded in favor of GBA's[16].
Going through the definitions for these ten synsets so far, one can notice three main aspects covered: the process of forming a rock (e.g.: consolidation, compaction, cementation); the constituents of such rock (e.g.: calcite, aragonite); and the size or aspect of the constituents (e.g.: rounded, $>$2mm). Focusing on the constituents, we confirm that *rock* is *substance meronym of* 14662574-n (mineral : solid homogeneous inorganic substances occurring in nature having a definite chemical composition) in WN. Reflectively, *mineral* is *substance holonym of rock*.
Combing through the definitions for the nine terms so far under *rock*, we see that the only minerals referenced are *calcite*, *aragonite* and *dolomite*. All three of them already exist in WN and required only minor changes in the definitions and/or the relations. Essentially the chemical formulas were added to the definitions and the *substance holonym of* relations according to the definitions of the terms we added to WN.
Finally, another set of relations was included: the *domain of synset - TOPIC* and *member of this domain - TOPIC*. Given our topic of choice, all of the terms we added from GBA's lithological terms were associated with *lithology* domain and their constituents with the *mineral* domain.
The *limestone* example shows our approach to map GBA into WN. We included six new and corrected four previously existent synsets definitions, along with their *hypernym of* and *hyponym of* relations. As we analyzed *substance holonym of* and *substance meronym of* relations, we included some of GBA's mineral terms in WN. It is not the scope of this work to cover all of GBA's minerals, but we included the ones mentioned in the rock's definitions.
Following this same approach, we were able to include all of *carbonate sedimentary rock* and *clastic sedimentary rock*, encompassing 27 new synsets with new 79 relations and 9 definitions changes, 15 removed relations and 71 new relations in pre-existing synsets.
These types of sedimentary rocks represent the two main types of oil & gas reservoirs throughout the world. By having them on WN we expect to move one step ahead in NLP for the Oil & Gas domain. We also expect that our time invested in ensuring proper synset relations will improve the performance of word sense disambiguation (WSD) algorithms, specially ones that rely on WN's graph such as UKB (Agirre and Soroa, 2009). At this point we ran the analysis covered in 7.. After the positive results, we carried on with our approach and finished the inclusion and mapping of all GBA lithology material into the WN. With this we expect to move one step further in NLP not only for the Oil & Gas domain but for all geological-related domains, such as Mining, Seismology, and so on.

6. Discussions

The extension of WN raised some relevant points. This section covers such points and explains the reasoning behind the decisions made within the possibilities considered.
A recurring matter regards the multiword expression (MWE) issue. Should we keep and create a synset for an MWE? Or is it enough to have all words individually in the resource? For instance, in WN we have 14698000-n (sedimentary rock : rock formed from consolidated clay sediments), but is it a 14696793-n (rock : material consisting of the aggregate of minerals like those making up the Earth's crust; "that mountain is solid rock"; "stone is abundant in New England and there are many quarries") that is 02952109-a (sedimentary : resembling or containing or formed by the accumulation of sediment; "sedimentary deposits")? Likewise, GBA subdivides *sandstone*, *sand*, *siltstone*, *silt* and *gravel* into *fine*, *medium* and *coarse*, meaning *fine* presents more and smaller grains than *medium* which in turn has more and smaller grains than *coarse*. But GBA sets a specific grain diameter range for *fine sandstone* which is different from the range of *fine siltstone* (respectively 0.063mm to 0.200mm and 0.0020mm to 0.0063mm). Due to this aspect, one possibility would be to adjust existing (or create new) synsets to ensure that *fine*, *medium* and *coarse* retain their relative properties, but the cutoff values (e.g.:0.063mm to 0.200mm) would be lost. In such cases we chose to respect our authoritative source.
Another issue we faced was when layman's knowledge

[14] Sedimentary material is defined in GBA as *a naturally-occurring material formed at the Earth's surface, consisting of solid particles aggregated together by one or more depositional processes operating within fluid systems (either aqueous or gaseous) to yield granular particles and/or crystalline particles that are aggregated into layers or bodies. The term includes both unconsolidated sediments and sedimentary rocks.*

[15] 14698000-n (sedimentary rock in WN : rock formed from consolidated clay sediments)

[16] Limestone definition in GBA is *A carbonate sedimentary rock composed of $>$ 95% calcite (and aragonite) and $<$ 5% dolomite*

clashes with technical definitions. For instance, 14698000-n (sedimentary rock : rock formed from consolidated clay sediments): from a technical perspective, clay is an unconsolidated sediment with very small grain, whilst *sedimentary rock* can be formed from several grain sizes, so we replaced WN's definition with GBA's. Another example is 14995541-n (sandstone : a sedimentary rock consisting of sand consolidated with some cement (clay or quartz etc.)). Even though WN's definition was not so far off, it presented *sandstone* as an *hyponym of* 14697485-n (arenaceous rock : a sedimentary rock composed of sand), a term not present in GBA. On the technical side *sand* is a clastic sediment within a certain grain size range, but on the other hand WN defines sand as being silica-based, i.e., the sand commonly found in beaches. This is a common misunderstanding even among technicians. In order to accommodate such divergent points, we merged *arenaceous rock* and *sandstone* synsets, kept the seven synsets *sandstone* was already *hypernym of* and then complemented with GBA's material.

7. Evaluation

In order to assess the impact of our project, we tested the same NLP pipeline in the same corpus once with the original PWN and once with our extended WN on its intermediary version, i.e. with only *carbonate sedimentary rock* and *clastic sedimentary rock* structures. The results confirmed the value of our approach and justified the inclusion of the remaining GBA's lithological terms.

The corpus used is one studied by (Rademaker, 2018). It consists of over five thousand sentences, with an average 28 words per sentence. It was built from 1298 publicly available English language geological reports, published by the United States Geological Survey, Geological Survey of Canada and British Geological Survey. The processing was done using Freeling 4.1 (Padró and Stanilovsky, 2012), with the corpus organized in one sentence per file.

The use of our OWN-EN implied in 910 words with different results. Nine had improper Part-of-Speech (PoS) tags and no sense attributed, and for those all PoS and senses were properly attributed with our OWN-EN, but only three to our new synsets - the other six were allocated to previously existing synsets. Such phenomena also happened where the PoS was already correct: of 78 words without allocated synsets, 69 were attributed to previous synsets and only 9 to new synsets. Another 184 words changed synsets within preexisting ones. Finally, there were 639 occurrences of *sandstone* that properly changed from the original WN synset to our previously discussed synset.

One interesting aspect that arises from such numbers is that, *sandstone* apart, most changes were to preexisting synsets. This shows the impact of adding and correcting relations within already existing synsets.

Another relevant case is the change from 13483488-n (formation : natural process that causes something to form; "the formation of gas in the intestine"; "the formation of crystal"; "the formation of pseudopods") to 09287968-n (formation : (geology) the geological features of the earth) for 59 occurrences of *formation*. Each case was checked, and the switch was judged appropriate for 51 of them. For the remaining eight cases the original synset was deemed correct.

Conglomerate has fourteen occurrences in the corpus, all of which were previously mapped to 08058937-n (conglomerate : a group of diverse companies under common ownership and run as a single organization) and afterwards were properly mapped to 14863031-n (conglomerate : a composite rock made up of particles of varying size). Each case was individually validated. To illustrate, an example sentence is presented below - clearly it is not about a group of companies, but rather composite rocks.

(1) On Pliocene and Pleistocene Siwalik Group fluvial sandstones and conglomerates mark the top of the stratigraphic column in the area

8. Conclusion

We were able to expand WordNet from an authoritative source, the Geological Survey of Austria Thesaurus (GBA). The process tackled with evaluating existing synsets for correctness when compared to GBA and creating new synsets otherwise. Such analysis comprehended not only definitions but also the conceptual relations that characterize WordNet.

A three step approach was used. We first used *limestone* as a proof of concept, then all of *carbonate sedimentary rock* and *clastic sedimentary rock*, the main types or reservoir rocks for Oil & Gas. The impact of such extension was evaluated with a corpus containing over five thousand sentences. The results indicated not only the relevance of new synsets added but also the impact conceptual relations changes have on old synsets. Finally, we extended WN to all of GBA's lithology.

Another product is the mapping between the extended WN synsets and GBA. Because GBA is also mapped to WikiData, BGS, Inspire, GeoSciML and DBpedia, our mapping links such resources as well. This mapping and the extended WN is available at `https://github.com/own-pt/own-en`.

9. Bibliographical References

Agirre, E. and Soroa, A. (2009). Personalizing pagerank for word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '09, pages 33–41, USA, 01. Association for Computational Linguistics.

Amaro, R. and Mendes, S. (2012). Towards merging common and technical lexicon wordnets. In *Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon*, pages 147–160, Mumbai, India, December. The COLING 2012 Organizing Committee.

Antoniak, M., Dalgliesh, J., Verkruyse, M., and Lo, J. (201 6). Natural language processing techniques on oil and gas drilling data. In *Intelligent Energy International Conference*, pages 1–6, September.

Bond, F. and Bond, A. (2019). Geonames wordnet (gnwn): extracting wordnets from geonames. In *Wordnet Conference*, page 387.

Buitelaar, P. and Sacaleanu, B. (2002). Extending synsets with medical terms. *Proceedings of the First International Conference on Global WordNet*, pages 21–25.

Encyclopaedia Britannica, T. E. o. (2018). *Limestone*. Encyclopædia Britannica, inc.

Fellbaum, C. (1998a). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press.

Christiane Fellbaum, editor. (1998b). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press.

Isaac, A. and Summers, E. (2009). Skos simple knowledge organization system primer. *Working Group Note, W3C*.

McCrae, J. P., Rademaker, A., Bond, F., Rudnicka, E., and Fellbaum, C. (2019). English wordnet 2019–an open-source wordnet for english. In *Wordnet Conference*, page 245.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1990). Wordnet: An on-line lexical database. *International Journal of Lexicography*, 3:235–244.

Muniz, H., Chalub, F., Rademaker, A., and de Paiva, V. (2018). Extending wordnet to geological times. In *Global Wordnet Conference 2018*, Singapore, January.

Padró, L. and Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. In *LREC2012*.

Palkowsky, B. (2005). A New Approach to Information Discovery - Geography Really Does Matter. In *SPE Annual Technical Conference*, pages 9–12, Dallas, October.

Parliament, E. and of the Council. (2007). Inspire directive 2007/2/ec. In *Official Journal of the European Union*, volume 50, page 371–es.

Peters, W., Sagri, M. T., Tiscornia, D., and Castagnoli, S. (2006). The lois project. In *LREC*, 01.

Poprat, M., Beisswanger, E., and Hahn, U. (2008). Building a biowordnet by using wordnet's data formats and wordnet's software infrastructure: A failure story. In *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, SETQA-NLP '08, page 31–39, USA. Association for Computational Linguistics.

Rademaker, A., Tessarollo, A., Pease, A., and Muniz, H. (2019). Extending sumo to geological times. In João Paulo A. Almeida, et al., editors, *Proceedings of the XII Seminar on Ontology Research in Brazil*, volume 2519, pages 70–82, Porto Alegre, RS, September. See http://ceur-ws.org/Vol-2519/.

Rademaker, A. (2018). Challenges for information extraction in the oil and gas domain. In Joel Luís Carbonera, et al., editors, *Proceedings of the XI Seminar on Ontology Research in Brazil (ONTOBRAS)*, São Paulo, Brazil.

Roventini, A., Antonietta, A., Bertagna, F., Calzolari, N., Jessica, C., Girardi, C., Magnini, B., Marinelli, R., Speranza, M., and Zampolli, A. (2003). Italwordnet: building a large semantic database for the automatic treatment of italian. *Linguistica computazionale : XVIII/XIX, 1998/1999*.

Sagri, M. T., Tiscornia, D., and Bertagna, F. (2004). Jurwordnet.

Schoen, E., Smith, R., and Boden, J. (2018). AI Supports Information Discovery and Analysis in an SPE Research Portal. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, September.

Smith, B. and Fellbaum, C. (2004). Medical wordnet: A new methodology for the construction and validation of information resources for consumer health. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, page 371–es, USA. Association for Computational Linguistics.

van Assem, M., Gangemi, A., and Schreiber, A. (2006). Rdf/owl representation of wordnet. Technical report, World-Wide Web Consortium W3C. https://www.w3.org/TR/wordnet-rdf/.

Vossen, P. (2002). Eurowordnet general document. eurowordnet (le2-4003, le4-8328), part a, final document.