Extending Wordnet to Geological Times

Henrique Muniz
EMAp/FGV, Brazil

Fabricio Chalub
IBM Research, Brazil

Alexandre Rademaker
IBM Research and EMap/FGV, Brazil

Valeria de Paiva
Nuance Communications, USA

Abstract

This paper describes work extending Princeton WordNet to the domain of geological texts, associated with the time periods of the geological eras of the Earth History. We intend this extension to be considered as an example for any other domain extension that we might want to pursue. To provide this extension, we first produce a textual version of Princeton WordNet. Then we map a fragment of the International Commission on Stratigraphy (ICS) ontologies to WordNet and create the appropriate new synsets. We check the extended ontology on a small corpus of sentences from Gas and Oil technical reports and realize that more work needs to be done, as we need new words, new senses and new compounds in our extended WordNet.

1 Introduction

Princeton WordNet (Fellbaum, 1998) works well as a dictionary and thesaurus for uses of English, as found, for instance, in newspapers and general knowledge texts, such as Wikipedia. Some attempts at extending it, for specific domains, such as Bioinformatics, Geography or Law (Smith and Fellbaum, 2004; Buscaldi and Rosso, 2008; Sagri et al., 2004; Lazari and Zarco-Tejada, 2012) have been made, but it is not clear how these extensions should be done, as different stakeholders will want to extend the basic dataset into different directions and with different tools and objectives.

The goal of our work is to describe a possible process of extension of the basic Princeton WordNet, for a restricted domain (Geological Time Periods) and to discuss issues, challenges and opportunities for other generic extensions.

One might wonder whether extensions of WordNet are really necessary. To this we say that even for everyday language we are convinced that WordNet misses some necessary synsets. For example, there are several issues related to tokenization: words like *ping-pong, kickboxing, water-ski* and *fistfight* should appear with space, hyphens or not, in the respective synsets. They do not, which means that quite a bit of post-processing is necessary. It would be good to add many prefixes, suffixes and regular endings, which are perfectly understandable by humans, but not so much by machines, for instance *shirtless* and *localizer, focalizer* are not in WordNet. Also many verbs ending in *-ize, ise* or *ify* are not present in PWN, while being in Wiktionary, for instance *adjectivise, Africanize* or *incentify, girlify*.

We might also want to discuss **why** the kinds of extension of WordNet we describe in this work are useful. We offer two complementary explanations. First we want to use WordNet as a sort of "lightweight" ontology. As discussed in (Bobrow et al., 2007; de Paiva, 2011) while full comprehensive ontologies like SUMO (Niles and Pease, 2001) or Cyc (Matuszek et al., 2006) would be best for reasoning formally with the information in texts, these tend to be very ragged. They only have detailed information in the specific domains that people felt the need to complete them for. For daily words and everyday, commonsense, events they miss many concepts. Some shallow reasoning can be done on the basis of the information provided by lexical taxonomies and it seems best to cover all concepts, at the expense of being shallow than to have big gaping holes in the concepts covered.

The second explanation has to do with bootstrapping specific domain ontologies for specific domains. Even if we did have a fully comprehensive version of an open source ontology for commonsense, we would still need to complement it for specific domains like Geology and Paleontology. There are too many concepts specific to the

field that English fluent speakers have never heard of and that should not be part of a basic lexical resource for English. But these specific, say, geological concepts, need to be fitted within the taxonomic framework of a lexical knowledge base like WordNet, so that we can take advantage of the aforementioned framework. Some of us would like to use this aspect of WordNet expansion to construct Gas and Oil ontologies for supporting projects on information extraction on that industry.

In the (small) experiments we report in this paper, we discuss a very specific extension to a hopefully not very controversial domain. We want to add to WordNet specific information concerning geological time periods. The geologic time scale (GTS) is a system of chronological dating that relates geological strata of rocks (stratigraphy) to time as measured in years in Earth's history.

2 Geological Time Periods

The geologic time scale is used by geologists, paleontologists, and other Earth scientists to describe the timing and relationships of events in Earth's history. The table of geologic time spans set forth by the International Commission on Stratigraphy, which we take to be the official body for these scientists, is described in `http://www.stratigraphy.org`.

Both Wikipedia and Wiktionary have some information about geologic time periods that seem more complete than the information in WordNet. This is to be expected, as lexicographers tend to be conservative about the terms they add to their repository of the language. But to be useful, when analyzing scientific texts about geological descriptions, we need to take the newer and more specific information present in the Wiktionary and Wikpedia in consideration. This is a common pattern. For several specific domains Wikipedia and Wiktionary have more current and more specific information than WordNet. WordNet is concerned about not inflating the lexicon with terms that are clearly derived, when looked from a human perspective, (e.g. *coaly* is simply the adjective form of having to deal with *coal*) or easily compositional (like *basinward*– in the direction of a basin). Also new expressions consisting of prefixes and suffixes are not considered good material for WordNet, so WordNet has *aeon*, but not *super-aeon*.

We would like to devise and describe a process to extend WordNet for a specific domain, when we do have specific information about the domain in the shape of a well curated ontology for the domain, as well as high quality texts in the same. We use geological time periods and a small collection of papers in Petrology as a paradigmatic example of a domain specific extension.

2.1 Geological Time in WordNet

Princeton WordNet has only 28 synsets dedicated to the most well-known geological periods. All the information about geological periods is concentrated on synsets that are hyponyms of [`15116283-n`: *geological time, geologic time - the time of the physical formation and development of the earth (especially prior to human history)*]. Hyponyms include synsets for each of *aeon, geological era, geological period* and *epoch*. We discuss briefly the essentials on these synsets below.

The geologic time scale is organized in a hierarchical fashion. Eons (or aeons) are divided into eras. Eras contain periods that contain epochs, and finally epochs contain ages. The first three eons (Hadean, Archean, Proterozoic) are collectively referred as the Precambrian super-eon. The most recent eon, the Phanerozoic is subdivided into several periods. All of these five names of periods have their respective synsets in WordNet, but *super-eon* is not in any synset. However, geologists and paleontologists need more detail than the 28 synsets in PWN provide.

The International Commission on Stratigraphy, a sub-comittee of the International Union of Geological Sciences, publishes regularly the International Chronostratigraphic Chart[1] as the current standard of the organization of the geologic timescale of the Earth. One can read about the development of the chart in (Cohen et al., 2013). As explained in that paper, geological time periods are not as well-established as one might expect. They say:

> Most of the systems, series and stages were first defined from type-sections in Europe, the historical home of stratigraphy. Subsequent study of stratigraphical successions worldwide has led to a proliferation of regional units. These histor-

[1] `http://www.stratigraphy.org/index.php/ics-chart-timescale`

> ical units did allow Phanerozoic strata to be correlated and mapped worldwide. However, as it happened, most successive chronostratigraphic units are located in geographically separated type sections, which have more recently been shown to be separated by significant gaps or to overlap considerably. These problems, and the general lack of defined boundaries for historically established units, became serious hindrances to high-resolution correlation of geographically widespread stratigraphic successions.

A committee was tasked with producing a chart that solved the issues of conflicting and overlapping regional strata. We assume the chart and the new periods and boundaries represent the consensus between scientists working on this area. The chart mentioned above contains 176 names of geological periods. Of these only 28 are in WordNet and all but 40 are in Wiktionary. The last 11 are in Wikipedia, but not in WordNet or Wiktionary.

While the common noun *stratigraphy* is in PWN, [`06118236-n`: *stratigraphy - the branch of geology that studies the arrangement and succession of strata*], even the adjective *stratigraphic* is not in the database and neither is the compound *chronostratigraphic*. Presumably because these words are too specific and their meaning can be easily derived from the prefix *chronos*, meaning 'time' and the suffix denoting a pertainym adjective *-ic*. However, even the word *strata* (the irregular plural of *stratum*) used in the gloss, and presumably more primitive than *stratigraphy* (the study of strata) is not in WordNet, which signals clearly that PWN needs to be extended, if it is to deal with the needs of the area.

One reasonable way of extending a lexical resource in the direction of a specific field is to process a corpus of quality texts in this field and check for missing entries. This was part of our work for this experiment. But another avenue of expansion open to us, in this case, was to incorporate a domain-specific ontology created by the professionals of the area. We searched for experts and found the ISC ontology `http://resource.geosciml.org/def/voc/`, described in the next section.

We should note though that the new ontology is not a full solution to our problem. There are many compounds and single words that acquire specific meanings within a field. Finding and creating synsets for these is also part of our challenge. Also, discovering when compounds are to be treated as multiword expressions, as opposed to compositional compounds, is a challenge, compounded by the use of abbreviations, specific to the field.

For instance, one of the main concepts of the area, the idea of a GSSP (Global Boundary Stratotype Section and Point [2]), is usually called a *golden spike* in text. Anyone who is not from the field might think that a golden spike is just a compositional English compound. Seeing the expression by itself, without context, they might not know that the expression stands for "an internationally agreed upon reference point on a stratigraphic section which defines the lower boundary of a stage on the geologic time scale", as explained.

We first discuss how to incorporate the information from an already structured ontology and then how to use corpora to improve our specific lexicon of geological time scales.

3 The ISC Ontology

The ISC ontology presents a view of the knowledge associated to the International Stratigraphic Chart. The ISC ontology contains many sub-ontologies, including the Geologic Timescale (GTS[3]) that would seem perfect for our uses.

In this ontology, *age, eon, epoch, era, period, sub-period*, and *super-eon* are sub-classes of `GeochronologicEra` (abbreviated as GE), which seems simply a different name for what is called 'geological time' in WordNet. However, there is no formally defined hierarchy between these concepts. Instead, greater emphasis is placed on the boundaries of the periods and only the approximate duration of the period is given in the chart. It is important to note that geologists qualify the units as "early", "mid", and "late" when referring to time, and "lower", "middle", and "upper" when referring to the corresponding rocks. For example, the lower Jurassic Series in chronostratigraphy corresponds to the early Jurassic Epoch in geochronology. The adjectives are

[2] `https://en.wikipedia.org/wiki/Global_Boundary_Stratotype_Section_and_Point`

[3] `http://resource.geosciml.org/ontology/timescale/gts.html`

capitalized when the subdivision is formally recognized, and lower case when not; thus "early Miocene" but "Early Jurassic".

While the commission was created exactly to unify and organize the classification of both strata and geochronological periods, it appears that the work is both not finished and bound to disagreement. The above mentioned paper also says

> [...] disagreement often arises, because type sections that are favoured for historical reasons may be abandoned, previously established boundary levels may be greatly changed, and in some instances historical units are replaced by different new ones.

Thus while the ontology might look very much a finished product, it seems that its contents are still subject to debate.

The boundaries between periods seem to be annotated using another ontology, the Temporal Hierarchical Ordinal Reference System model (THORS[4]), which is used to formally define the hierarchy between instances of GE. Fragments of the `ISO19108:2002` standard (Geographic information – temporal schema) are also used to specify the temporal position of geochronologic boundaries.

The time interval of a GE is given in terms of its boundaries to other GEs via `thors:begin` and `thors:end`. Each boundary is a `GeochronologicBoundary` and it is temporally located via `iso19108:temporalPosition` which specifies a `iso19108:Coordinate` with a value, frame (e.g., "Ma"), and a positional uncertainty.

For example, the Maastrichtian period is defined by Wiktionary in `https://en.wiktionary.org/wiki/Maastrichtian` as "in the ICS geologic timescale, the latest age or upper stage of the Late Cretaceous epoch or Upper Cretaceous series, the Cretaceous period or system, and of the Mesozoic era or erathem".

In the ISC ontology itself the definition is more complex. The Maastrichtian period (66–72.1 Million years) is defined using boundaries and frames (Figure 1).

[4] `http://resource.geosciml.org/ontology/timescale/thors.html`

```
Maastrichtian a GeochronologicEra ;
  rank Age ;
  begin BaseMaastrichtian ;
  end BaseCenozoic .
BaseMaastrichtian a GeochronologicBoundary ;
  temporalPosition BaseMaastrichtianTime .
BaseCenozoic a GeochronologicBoundary ;
  temporalPosition BaseCenozoicTime .
BaseMaastrichtianTime a Coordinate ;
  frame ma ;
  value "72.1" .
BaseCenozoicTime a Coordinate ;
  frame ma ;
  value "66" .
```

Figure 1: A fragment of the Maastrichtian period definition on ISC ontology

The boundary modeling should be sufficient for representing the hierarchical relationship between GEs, but ISC further defines a explicit set inclusion relationship between GEs via the `thors:member` property. Also, SKOS (Isaac and Summers, 2008) is also used to represent inclusion via `skos:narrower`, `skos:broader` along with theirs transitive versions, `skos:narrowerTransitive` and `skos:broaderTransitive`.

In any case a collection of 176 basic geologic period terms is easy to deal with, if the scientists are in agreement. However, we still have to deal with common nouns (e.g. *play, basin, cleats*) and compounds (e.g. *golden spike*), whose geological meanings are very different from their usual meanings. These need to be extracted from a geology corpus, similar to the one we describe in the next section.

4 A corpus of Geological Reports

The source documents for the our small experiment come from 155 randomly selected text passages relevant to petroleum systems extracted from a corpus of 1,298 publicly available English language geological reports, published by the United States Geological Survey (USGS), Geological Survey of Canada (GSC), and British Geological Survey (BGS).

The passages were segmented in 5,661 sentences (186,244 tokens) and parsed in the Universal Dependencies scheme by UDpipe (Straka and Straková, 2017) [5]. UDpipe is a generic, off the shelf processing pipeline trained with the English corpus from the Universal Dependencies project

[5] `http://ufal.mff.cuni.cz/udpipe`

(Nivre et al., 2016). Using the model available, trained on newswire data, it does not do well on Named Entity recognition in our corpus. Our preliminary semantic pipeline looks up nouns, verbs, adjectives and adverbs in Princeton WordNet. Out of 8800 noun lemmas uncovered by UDpipe, more than half were not recognized as present in WordNet. Because the reports are describing real world geological work, the corpus is full of named entities, e.g. names of places, people and organizations that cause Named Entity Recognition to be such a hard task.

Some of these missing words are processing mistakes. For instance, the word 'reservoirs' was not correctly lematized to 'reservoir'. A large proportion are named entities, people, places and organizations that WordNet is not supposed to list in any case. But a small proportion are really common words that WordNet should have, in our opinion. Finding these seems to be a positive side effect of trying to extend WordNet for specific domains.

Since our aim is not the processing of this corpus, but simply its use as a source of extra vocabulary for our extended WordNet, we decided to look at all tokens in the corpus with more than 10 occurrences, trying to decide whether they were Named Entities or not. And we assumed that the processing could be corrected, by hand, if need be. It is well-known that PWN lacks some important compounds and that the cut-off line for compounds to be lexicalized is a difficult one to decide on. Moreover, in this specific field, we do not know exactly when compounds are compositional or not. But a shallow processing of the text provides us with some 20K proper nouns, so almost 4 proper nouns per sentence. This means that NER is a very hard job, even assuming near perfect Geoname resources, which unfortunately we do not have.

5 Creating New Synsets

The language of ISC and its various ontologies is complex, and for a reason. They want to be precise, while trying to merge different standards. As we want to map all their precision into an extended version of Princeton WordNet we need a kind of a *domain specific language* (DSL) to describe new synsets. This language helps us not only to describe the new synsets we need, but also should helps us localize these new synsets within the original WordNet structure.

The file format we decided to use is intended mainly for human consumption, even at the cost of a more complicated parsing routine. Redundancies are eliminated, for example there is no need to specify both sides of reflexive relations, such as hyponymy and hyperonimy. Artificial identities (ids) are avoided to make maintenance easy. Actual ids are based on the lexical units, following the ideas of the original lexicographer files for Princeton WordNet. Instead of using symbols such as `@`, `!`, etc. for relations, we use mnemonics such as `hyper` (hypernym) and `ant` (antonym). The goal is to make a standalone domain specific language – one that is usable without any accompanying integrated development environment (IDE) or other auxiliary program.

Synsets are defined by groups of lines, separated by a single empty line. Words of the synsets should have their spaces converted to underscores and repeated words in the same file should have suffixes to distinguish them, also following the original lexicographer files of PWN. For example the synset for *eon* will be written as

```
w: eon drf adj.pert:eonian
w: aeon drf adj.pert:aeonian
hyper: geological_time
g: the longest division of
geological time
```

where `drf` stands for 'derived form', `adj.pert` is the usual WordNet description of the pertainym adjective file and `g` stands for the 'gloss'. Each word entry is essentially a *sense*. Links between senses are specified in the same line as the `w:` word, for example:

```
w: uptime ant downtime
```

means that 'uptime' and 'downtime' are antonyms. Semantic relations (i.e., links between synsets) are specified on lines of their own, such as the hypernym `hyper: geological_time` above.

The first word of a synset is used as its identifier. The lexicographer file filename should also be included to further disambiguate words, if necessary. This is usually the case when there are semantic links across synsets defined in different files. For example, the file `noun.location` contains the following excerpt for the synset "Brazil":

```
w: Brazil drf adj.pert:Brazilian
hp: noun.object:South_America
```

To maintain compatibility with existing systems that already use PWN sense keys and synset ids we provide mappings between our sense ids and PWN. Similarly, mappings that link synsets and existing ontologies can also be defined.

The full set of PWN synsets extended with the nodes created for the geological time periods and the new concepts we deem necessary to understand our corpus could be called PWN$_{GTS}$ for WordNet extended for the Geological Time Scale. In the next section we describe a toy application of the extension developed. In `http://github.com/own-pt/wordnet-dsl` we provide the PWN$_{GTS}$ and the mappings from the new synsets to the ISC Ontology.

6 Using PWN$_{GTS}$

The following discussion showcases an example where a number of geochronologic entities may be referenced implicitly by the text. Consider the following sentence from our corpus:

> In this chapter, the kinematic interpretation of the west Carbonate shear zone is placed in a regional context, with regard to intrusive and tectonic activity from 2740 to 2690 Ma ago.

Assuming that a parser correctly identifies the numerical range above as being 2740–2690 and the unit 'Ma' (for a million years), one can use our extended WordNet, creating a query to the ISC ontology that searches for entities that encompass this period of time. The SPARQL query used is in the appendix, note that such a query does not take into consideration the variance of the boundaries of time periods (modeled by the ontology). We opted to omit this feature to keep the SPARQL code simple. This natural query is not enough to uniquely disambiguate the appropriate instance that is referenced above, since the query returns three ISC entries: the Neoarchean era (2500–2800 Ma), the Archean eon (2500–4000 Ma), and also the Precambrian super-eon (541–4567 Ma).

While this toy example shows one possible use we envisage for very restricted forms of extension of the basic English WordNet, the larger question of evaluating such extensions beckons. From our preliminary work we can see some possibilities, which we discuss next.

7 Evaluating Extensions

It is clear that different kinds of text and different content domains play a big role in the vocabulary that lexical resources are expected to cope with. This is clear for specific content domains, such as BioInformatics, where changes are recent and newer vocabulary is being created at impressive speeds. But even for domains, such as Geology, where one might have expected the main vocabulary to have been established by the end of the 19th century, things are not as well settled as expected.

Certainly there is a need for more (open source, downloadable) online glossaries, apart from the (small) Wikipedia one[6], the OpenLearn project[7] and the one from USGS[8] that has not been updated since the mid 2000's. But it seems that the proprietary ones still have the upper hand. The American Geosciences Institute (AGI) offers their fifth revised edition of the Glossary of Geology (Neuendorf, 2005) as a book and as paid subscribing content online. They say that their reference tool contains nearly 40,000 entries, including 3,600 new terms and nearly 13,000 entries with revised definitions from the previous edition. None of the open source glossaries we found has as many entries as that.

One way of measuring how much we can do with the open resources online is to measure how much of the informational contents of technical reports can be gleaned by a impoverished NLP pipeline that builds bag-of-concept semantics from the sentences of the chosen corpus. In a previous experiment we have computed this kind of bag-of-concepts semantics for sentences of the corpus SICK (Marelli et al., 2014). The corpus SICK is much easier to deal with, as it was engineered to not have any named entities at all. If we had no named entities in our geological reports, we could produce concepts from SUMO (Niles and Pease, 2001) using a bare bones pipeline that transforms sentences into universal dependencies (using UDPipe), dependencies into WordNet concepts or synsets (using, say, Freeling/UKB (Agirre and Soroa, 2009) for disambiguation) and WordNet synsets into SUMO con-

[6] `https://en.wikipedia.org/wiki/Glossary_of_geology`

[7] `http://www.open.edu/openlearn/science-maths-technology/science/geology/geological-glossary`

[8] `https://geomaps.wr.usgs.gov/parks/misc/glossarya.html`

cepts (using the SUMO mappings). An example of a processed sentence is displayed in Figure 2.

The idea here is not to produce knowledge representations of the meanings of the sentences, but simply to list the expressions for which we do not have a concept. For these 'empty concepts' we need either geographical information or new synsets, as they correspond to either new content words (that never appeared in WordNet before, like e.g. *vitrinite* or *stratigraphic*), or new compounds (e.g. *pre-Mississipian, antiform* or *subbasin, golden spike*) or new senses of words already in WordNet (e.g. *cleats, play, sequence*, which have completely different meanings in Geology from their usual ones). However we need to find a way of coping programmatically with named entities, for this baseline calculation to work.

Given the hardness of the NER problems in this particular kind of texts, we resorted to different open systems (with different training data and heuristics, e.g. OpenNLP [9] and Freeling (Padro and Stanilovsky, 2012)) to try to extract most of the named entities. In this corpus apart from locations, people and entities we have many *Geological Formations*, which span counties and even states' lines. To help debug our processing, we are experimenting with interfaces that allow linguists, computer scientists and geologists to communicate more easily `http://wnpt.brlcloud.com/demo`. We hope to improve, using subject matter experts, the number of new synsets and new senses. The manual 'ensemble' effort to recognize named entities we produced for this small corpus, needs to be streamlined in the future, for the work in extending other domains.

8 Conclusions

This preliminary work discusses extensions of Princeton WordNet for specific content domains. The case we considered is the well delimited domain of geological time periods. We expected it to be less controversial and to have a more established vocabulary than it turned out to have. However, we stand by our initial suggestion that specific domains require specific extensions. That these specific extensions need to be built as much as possible from open source resources, in a collaborative fashion, using as much as possible associated ontologies produced by the subject matter experts. However, a useful way to augment the specific knowledge required is to shallow process scientific texts on the specific subject (we used gas and oil technical reports) and try to extract more lexical information from them. Our small experiment with geological reports indicate that a more robust mapping of named entities is required before we can evaluate the usefulness of our new Geological Time Scale WordNet. We are working on a tool that would pre-annotate some of these geonamed entities and would facilitate the correction of the mistaken annotations.

[9] `https://opennlp.apache.org/`

References

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.

Daniel G Bobrow, Bob Cheslow, Cleo Condoravdi, Lauri Karttunen, Tracy Holloway King, Rowan Nairn, and Valeria de Paiva. 2007. PARCs bridge and question answering system. *Grammar Engineering Across Frameworks*, pages 46–66.

Davide Buscaldi and Paolo Rosso. 2008. Using geowordnet for geographical information retrieval. In *Evaluating Systems for Multilingual and Multimodal Information Access, 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers*, pages 863–866.

K.M. Cohen, S.C. Finney, P.L. Gibbard, and J-X Fan. 2013. The ICS International Chronostratigraphic Chart. *Episodes*, 36(3):199–204.

Valeria de Paiva. 2011. Bridges from language to logic: Concepts, contexts and ontologies. *Electronic Notes in Theoretical Computer Science*, 269:83–94.

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

Antoine Isaac and Ed Summers. 2008. Skos primer. Technical report, W3C. latest version available at http://www.w3.org/TR/skos-primer.

Antonio Lazari and M[a] Ángeles Zarco-Tejada. 2012. Jurwordnet and framenet approaches to meaning representation: a legal case study. In *Semantic Processing of Legal Texts (SPLeT-2012) Workshop Programme*, page 21.

Marco Marelli, Stefano Menini, Marco Baroni, Luisa Bentivogli, Raffaella Bernardi, and Roberto Zamparelli. 2014. A SICK cure for the evaluation of compositional distributional semantic models. In *Proceedings of LREC 2014*.

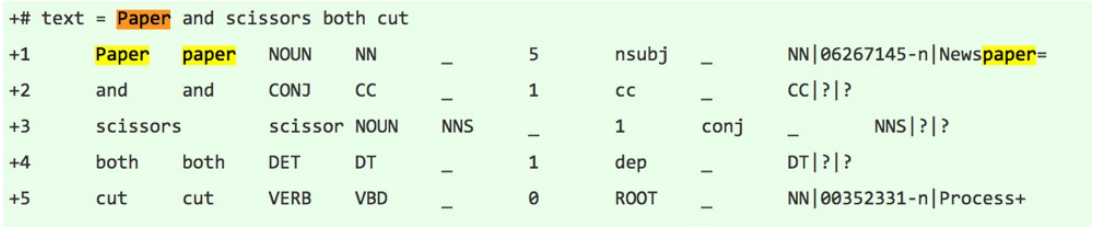

Figure 2: Bag-of-concepts

Cynthia Matuszek, John Cabral, Michael J Witbrock, and John DeOliveira. 2006. An introduction to the syntax and content of cyc. In *AAAI Spring Symposium: Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering*, pages 44–49.

Klaus KE Neuendorf. 2005. *Glossary of Geology*. Springer Science & Business Media.

Ian Niles and Adam Pease. 2001. Toward a Standard Upper Ontology. In Chris Welty and Barry Smith, editors, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, pages 2–9.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan T McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *LREC*.

Lluis Padro and Evgeny Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, May. ELRA.

Maria Teresa Sagri, Daniela Tiscornia, and Francesca Bertagna. 2004. Jur-WordNet. In Petr Sojka, Karel Pala, Pavel Smrz, Christiane Fellbaum, and Piek Vossen, editors, *Global Wordnet Conference*.

Barry Smith and Christiane Fellbaum. 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, Stroudsburg, PA, USA. Association for Computational Linguistics.

Milan Straka and Jana Straková. 2017. Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada, August. Association for Computational Linguistics.

A Example of Query

In the query below, notice if we remove the restriction on `isc:rank Age` we get multiple hits (Maaast. [age], Cret. [period], Upper Cret. [epoch]) since the range 67–70 is included on all of them.

```
select ?era ?rank ?vbegin ?vend
{
 ?era gts:rank ?rank ;
   thors:begin ?tb;
   thors:end ?te .

  ?tb ts:temporalPosition ?begin;
  ?te ts:temporalPosition ?end .

  ?begin ts:frame age:ma ;
     ts:value ?vbegin .

  ?end ts:frame age:ma ;
     ts:value ?vend .

  bind (2690 as ?a)
  bind (2740 as ?b)

  filter ((?a <= ?vbegin &&
           ?a >= ?vend) ||
          (?b <= ?vbegin &&
           ?b >= ?vend))
}
```