A Semantic Role Repository Linking FrameNet and WordNet

Irina Sergienya

sergienya@gmail.com

University of Trento, Italy

Volha Bryl, Sara Tonelli

{bryl,satonelli}@fbk.eu

Fondazione Bruno Kessler, Trento, Italy

1 Creation of the Repository

FrameNet [4] is one of the most important semantic resources encoding information about situations, the *frames*, and participants, the semantic roles, also called *frame elements* (FEs). FrameNet is currently characterized by an extremely high number of roles, which amount to 8,884 in the last resource release (version 1.5). In order to provide a better generalization over all these roles, around 40 semantic types have been defined by FrameNet lexicographers to provide semantic constraints on FE fillers (e.g. the semantic type *Sentient* assigned to the *Agent* FE). Still, these semantic types cover only 54% of all FEs.

In order to tackle the problem of role generalization and provide a semantic characterization of the typical role fillers, we automatically created a mapping between FrameNet roles and WordNet synsets [1]. The resource, which we refer to as *sense repository*, was created in a bottom-up fashion, starting from the FrameNet corpus, so that we were able to assign a (statistically interpretable) weight to each synset associated with a FrameNet role (for details on the workflow, see also [5]). After each role filler was disambiguated by assigning a WordNet synset to it, a further generalization was performed by selecting from WordNet taxonomy [2] only the synsets that most frequently dominated the roles fillers.

Let us give an example of an entry of the repository. For the *Entity* semantic role of the AGING frame, 38 annotated examples can be found in the FrameNet corpus. After disambiguating the role fillers and generalizing through WordNet taxonomy, the following two lines are added to AGING–*Entity* file of the repository:

```
person_100007846, 36, 0.947368421052632

equipment_103294048, 2, 0.0526315789473684
```

suggesting that around 95% of the examples of the role fillers fall in *person* and 5% in *equipment* category, where the category is a WordNet synset (the 9-digit number in a WordNet synset name is its numerical ID).

While the first version of the sense repository was created in plain text we further converted it into RDF/OWL in order to make it available to the Semantic Web.

2 Conversion into RDF

We have modeled the structure of the repository as an ontology. Specifically, the main class of the ontology is `learntSemType`, each individual of which corresponds to one line in the repository and represents one possible category for a filler for a particular frame-role pair. For example, for the AGING frame and for the *Entity* role, two entities of `learntSemType` were created: one for *person* WordNet synset (`lst-Aging-Entity-person_100007846`), and one for *equipment* WordNet synset (`lst-Aging-Entity-equipment_103294048`).

We used WordNet 3.0 RDF[1] representation to connect our resource to WordNet. For the structure of the OWL version of FrameNet 1.5, we rely on [3]. However, as the populated ontology is not yet available, we used FrameNet 1.5 XML representation as a reference for defining frame and semantic role URIs within the ontology.

The properties of an individual of `learntSemType` class and the schematic relations between FrameNet, WordNet and our resource are depicted in Figure 1.
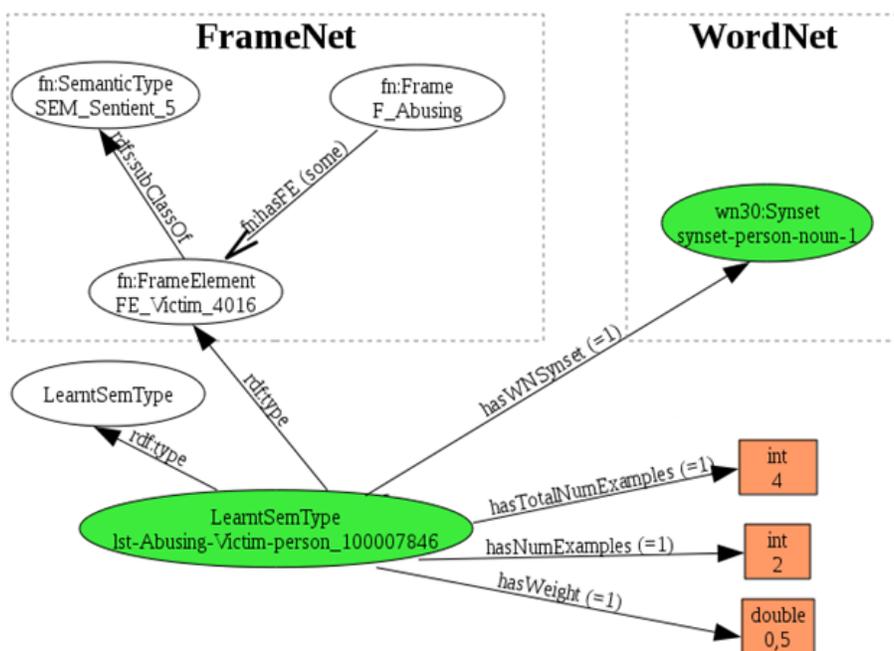

Figure 1: Structure of the resource and its relations to WordNet and FrameNet.

As can be seen from Figure 1, `learntSemType` is a subclass of the *FrameElement* class, which is a class in the OWL representation of FrameNet. For example, `lst-Abusing-Victim-person_100007846` is an individual of the class `FE_Victim_4016`. In addition, each individual of the `learntSemType` class has the following properties:

- `hasWNSynset` links an individual to a WordNet 3.0 synset;

---

[1]`http://semanticweb.cs.vu.nl/lod/wn30/`

- `hasTotalNumExamples` links an individual to the number of examples in the corpus annotated with the frame – FE this individual corresponds to (e.g. there are 4 examples in the corpus annotated with `FE_Victim_4016`);

- `hasNumExamples` links an individual to the number of examples in the corpus annotated with the frame – FE pair this individual corresponds to, that are classified with WordNet synset or its hyponym specified by `hasWNSynset` property (e.g. 2 examples annotated with `FE_Victim_4016` were classified either as `person_100007846` or its hyponym);

- `hasWeight` links an individual to the weight associated with (frame, FE, WordNet synset), which in the simplest case is the rate `hasNumExamples/hasTotalNumExamples` (for more details see [5]).

A URI of an individual of the ontology is in the following format:

`PATH/Frame#lst-Frame-FrameElement-WNSynset,`

for example, `PATH/Abusing#lst-Abusing-Victim-person_100007846,`

where `PATH` is `https://dkm.fbk.eu/FrameNet_SenseRepos/senses`.

Such a URI resolves in document that contains RDF representation of all individuals corresponding to a given frame.

The OWL version of the resource is downloadable from
`https://dkm.fbk.eu/index.php/FrameNet_extension:_repository_of_senses`
together with the tools for quering, filtering and counting some statistics on the repository.

The resource is also registered with CKAN:
`http://thedatahub.org/dataset/framenet-sense-repository`.

The RDF version of the sense repository consists of **24,569** individuals (corresponding to objects of `learntSemType` class). Each individual is connected to FrameNet and WordNet, and contains statistical information through `hasTotalNumExamples`, `hasNumExamples` and `hasWeight properties`. Therefore, the resource contains **122,845** meaningful triples[2].

---

[2]Current version of the resource contains more triples, e.g. to preserve the ontology structure or to facilitate the query processing; as this is a matter of further optimization, we do not include these triples in the number reported above.

References

[1] V. Bryl, S. Tonelli, C. Giuliano, and L. Serafini. A Novel FrameNet-based Resource for the Semantic Web. In *Proceedings of ACM Symposium on Applied Computing, Technical Track on The Semantic Web and Applications*, Riva del Garda, Trento, 2012.

[2] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database.* MIT Press, 1998.

[3] A. G. Nuzzolese, A. Gangemi, and V. Presutti. Gathering lexical linked data and knowledge patterns from framenet. In *Proceedings of the sixth international conference on Knowledge capture*, pages 41–48, 2011.

[4] J. Ruppenhofer, M. Ellsworth, M. R. Petruck, C. R. Johnson, and J. Scheffczyk. *FrameNet II: Extended Theory and Practice.* Available at http://framenet.icsi.berkeley.edu/book/book.html, 2010.

[5] S. Tonelli, V. Bryl, C. Giuliano, and L. Serafini. Investigating the Semantics of Frame Elements. In *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (to appear)*, Galway, Ireland, 2012.