

# Search tool for An Event-Type Ontology

Nataliia Petliak, Cristina F. Alcaina, Eva Fučíková, Jan Hajič, Zdeňka Urešová

ÚFAL MFF UK

Charles University, Faculty of Mathematics and Physics, Computer Science School

Malostranské nám. 25, Prague 1, Czech Republic

{petliak,alcaina,fucikova,hajic,uresova}@ufal.mff.cuni.cz

## Abstract

This short demo description paper presents a new tool designed for searching an event-type ontology with rich information, demonstrated on the SynSemClass (Urešová et al., 2022) ontology resource (version 5.0, (Urešová et al., 2023)). The tool complements a web browser, created by the authors of the SynSemClass ontology previously. Due to the complexity of the resource, the search tool offers possibilities both for a linguistically-oriented researcher as well as teams working with the resource from a technical point of view, such as building role labeling tools, automatic annotation tools, etc.

**Keywords:** language resource, lexical semantics, ontology, event types, demonstration, search tools, user interface

## 1. Introduction

Attempts aiming at improving formalized knowledge representation have resulted in the development of a number of huge lexical databases. Some of them are being interconnected to facilitate cross-formalism or cross-language studies. An exemplary project of this kind is a well-known initiative connecting lexical semantic resources called SemLink (Stowe et al., 2021), which aims to link together different mostly verb-oriented lexical semantic resources via a set of mappings; specifically, mappings between different word senses and semantic roles, as well as annotated corpus data. SemLink uses an online presentation system called The Unified Verb Index (UVI)<sup>1</sup> which merges those links and web pages from five different NLP projects.

All ontologies mentioned in SemLink are in one way or another related to the SynSemClass ontology for which the search tool described in this article has been created and on which it has been tested. However, the character of the search engine described herein allows it to be applied to other event-type ontologies as well.

Conversely, the search tool has been inspired in part by the UVI index search and other tools as available e.g., in BabelNet (Navigli and Ponzetto, 2010, 2012).

The organization of this paper is as follows: Sect. 2 introduces the SynSemClass ontology. Sect. 3 describes the already existing web browser used for browsing the ontology. Sect. 4 describes the core of the paper, the search engine, its interface, and examples to demonstrate its capabilities. We conclude and draw future plans in Sect. 5.

<sup>1</sup><https://uvi.colorado.edu/>

## 2. The SynSemClass Ontology

For exemplifying the search tool, we have used the SynSemClass ontology.<sup>2</sup> There also exists a web-based browsing tool.<sup>3</sup> SynSemClass can be downloaded in full as a set of XML files.<sup>4</sup>

We present here a short description of the resource, taken from (Urešová et al., 2020). SynSemClass is one of the lexical semantic oriented projects dealing with the most common form of expressing events and states, namely verbs, and semantic role labeling.

SynSemClass concentrates on the participants of these events or states and the relations between them. For these relations the term “semantic roles” is used. Unlike other resources representing verb semantics, such as PropBank(s) (Kingsbury and Palmer, 2002), WordNet(s) (Fellbaum, 1998), FrameNet(s) (Baker et al., 1998), or VerbNet (Bonial et al., 2012), SynSemClass has been designed from the start as an “inter-lingual” resource, representing cross-lingual meaning of verbs (currently) in English, Czech, German and Spanish, including links to 18 other external lexical-semantic resources in these languages. Furthermore, SSC maps the valency behavior of the class members to the set of semantic roles in each class (for more details see (Urešová et al., 2020)). The SynSemClass classes are meant to represent eventive concepts “universally,” i.e., to multiple, typologically diverse languages in a single resource.

SynSemClass in its current version 5.0 includes 1,546 classes with 15,790 Class Members (lexical units). All classes have Czech and English mem-

<sup>2</sup><https://ufal.mff.cuni.cz/synsemclass>

<sup>3</sup><https://lindat.cz/services/SynSemClass50/>

<sup>4</sup><http://hdl.handle.net/11234/1-5230>

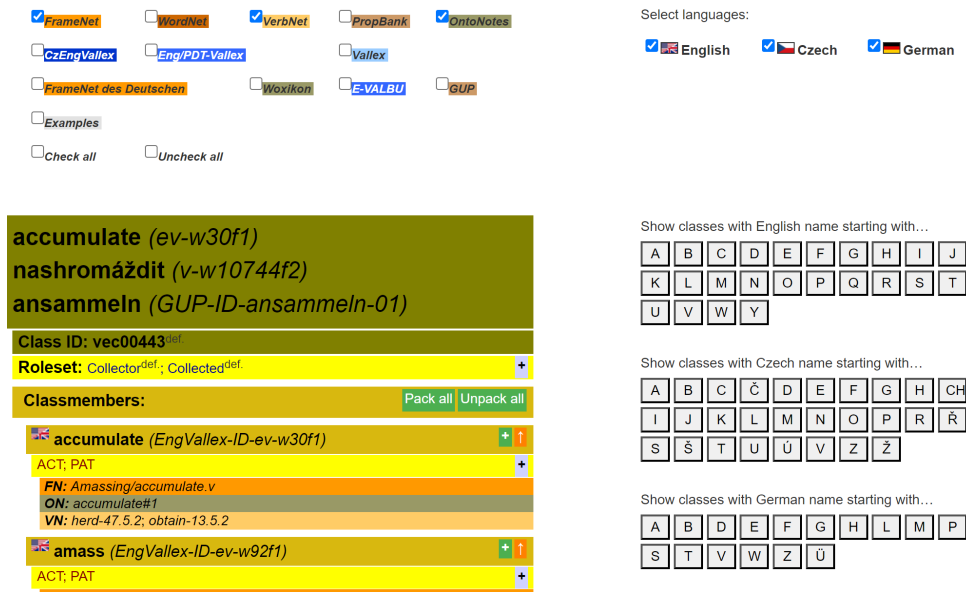


Figure 1: SSC web browser: (partial) display of results (*accumulate* class) as selected in the checkboxes

bers, and some of them also German and Spanish ones; the authors plan to extend it to more languages (Urešová et al., 2022).

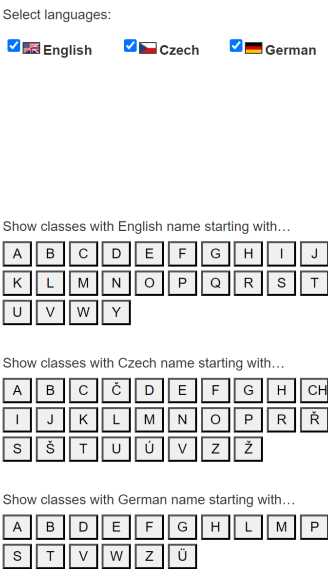
### 3. The SynSemClass Web Browser

The existing web browser offers limited options for finding entries (it is truly just a browser), but it enables to selectively display all information contained in the ontology. Thus the search tool has been designed to make use of the browser for a fully formatted display of the search results based on user's post-filtering of the search results.

The web browser interface is divided into two parts: the contents are displayed on the left while the languages and entries to select from are on the right. The default view shows all contents, including external links, for all available languages. The user can select the resources and/or languages desired for browsing by ticking the box located in front of each resource label (Figure 1).

The entries to display are selected through the list(s) under "Show classes with [chosen language] name starting with..." The user first selects a letter and then a class by its name starting by that letter.

Additional information can be accessed by hovering over specific items. E.g., definitions of the class and the roles defined in the Roleset are displayed in a pop-up window when the mouse is over the superscript "def" located to the right of the Class ID and name of the role, respectively.



## 4. The Search Tool

### 4.1. The SSC Search Tool

The search tool<sup>5</sup> offers multiple search criteria and flexible functionality for combining search options and building complex queries. The server-side development of the tool utilizes Express.js, which is retrieving data from the MongoDB database of the (converted) ontology data, instead of accessing the XML files directly. A React application is used on the client side.

Figure 2 depicts the tool's overall appearance.

### 4.2. Search Fields and Logic

The search interface contains the following search fields and query builders:

- **Lemma** - a class member's lemma, e.g., "bring".
- **Sense ID** - the sense ID of the class member, e.g., "EngVallex-ID-ev-w122f2".
- **Class ID** - common class ID to which the class members belong, e.g., "vec00107".
- **Filters** define the search languages. SynSemClass 5.0 currently includes English, Czech, German and Spanish, but the ontology will be expanded to include additional languages in the future. By default, all languages are searched; the user may select one or more languages to limit the search to those languages.

<sup>5</sup><https://lindat.mff.cuni.cz/services/SynSemClassSearch/>; for the project description, API description and guidelines and general context see also <https://ufal.mff.cuni.cz/synsemclass/synsemclass-search-tool>

Filter languages  English  Czech  German

Lemma search:  Sense ID search:  Class ID search:

Role(s) search:

Construct roles query in Conjunctive Normal Form,  
 e.g., (Role1 OR Role2) AND (Role3 OR Role4) AND (Role5)

Select the role(s), drag and drop selected roles into the clauses brackets or between them.  
 For mobile view, select the role; if needed, add more clauses and select more roles within each clause.

Figure 2: Search tool

Filter languages  English  Czech  German

Lemma search:  Sense ID search:  Class ID search:

Role(s) search:

Construct roles query in Conjunctive Normal Form,  
 e.g., (Role1 OR Role2) AND (Role3 OR Role4) AND (Role5)

Select the role(s), drag and drop selected roles into the clauses brackets or between them.  
 For mobile view, select the role; if needed, add more clauses and select more roles within each clause.

Figure 3: Sample query: desktop view

- **The roles section** defines the search based on the roles of class members. The user can select a role from a list of available options. Conjunctive Normal Form (CNF) is used to perform a more advanced search of multiple role combinations. The user can manipulate the brackets (CNF clauses) by inserting roles between them and by adding additional clauses. The final query regarding roles is displayed at the bottom. It is always in CNF format, such as (Role1 OR Role2) AND (Role3). The desktop version of the web employs a drag-and-drop user interface for interactive and visual query creation. The mobile version of the application retains the capability to expand the roles query with additional drop-down options for each added clause.

Regular expressions can be used in any input field to match any strings. For example, “put.\*” will show class members starting with “put”.

A sample query combining multiple search criteria is shown in Figures 3 and 4.

### 4.3. Presentation of Results

The summary of the results includes the number of class members and unique classes found in total, as well as by language (Figure 5).

The data representation is class-centric. Class members matching the criteria are grouped into their common classes, with essential information condensed at the top level.

The class ID is displayed at the top. The roles

are highlighted in green. Within a given class, only the top two class members with their respective sense IDs are displayed showing collapsed state of the result (Figure 6).

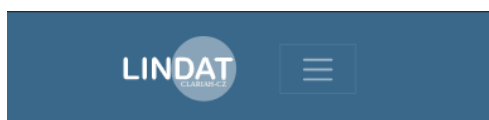
The “Show more” button reveals the complete list of class members with their sense ID, class member ID, and mapping (Figure 7).

Finally, expanding each individual class member with the arrow on the right reveals the member’s complete JSON content (Figure 8).

Both on the top-level presentation of results and inside fully expanded JSON content, the data contains links to external sources. For instance, clicking on a class ID or class name would highlight a link pointing to the corresponding class in the SynSemClass web browser. Similarly, clicking on the class member’s lemma would lead directly to a class member in the browser. In addition, in the fully expanded JSON output, highlighted fields also contain links to the corresponding entry in the external lexicon browsers (PDT-Vallex, EngVallex, etc.). For user convenience, class ID, sense ID and class member ID fields reveal a copy icon upon mouseover, allowing for a quick copy of the contents of these fields (Figure 9).

### 4.4. Search Tool API

The application was developed according to RESTful API specifications. By utilizing API endpoints, users can send search queries and receive rendered results identical to those obtained through the UI, or they can directly receive the raw response from the server for further processing



Filter languages  English  Czech  German

Lemma search:

Sense ID search:

Class ID search:

Role(s) search:

Construct roles query in Conjunctive Normal Form,

e.g., (Role1 OR Role2) AND (Role3 OR Role4) AND (Role5)

Select the role(s), drag and drop selected roles into the clauses brackets or between them.

For mobile view, select the role; if needed, add more clauses and select more roles within each clause.

Objection x

( Agent x OR Recipient x ) x

AND

( Offered x ) x

**(Agent OR Recipient) AND (Offered)**

Figure 4: Sample query: mobile view

**Found 9 class member(s) in 7 unique class(es).**

- 9 English class member(s) in 7 class(es)
- 0 Czech class member(s) in 0 class(es)
- 0 German class member(s) in 0 class(es)

Figure 5: Search results information summary

**class ID: vec00014**

**roles: Agent, Entity, State**

**class: keep (ev-w1792f2)**

7 class member(s):

have (EngVallex-ID-ev-w1566f16\_u\_nobody)

hold (EngVallex-ID-ev-w1601f14\_u\_nobody)

...

Figure 6: Initial presentation of the result

**class ID: vec00014**

**roles: Agent, Entity, State**

**class: keep (ev-w1792f2)**

7 class member(s):

have (EngVallex-ID-ev-w1566f16\_u\_nobody) (vec00014-eng-cm00001) ACT ... Agent, PAT ... Entity, #alt[MANNLOC] ... State

hold (EngVallex-ID-ev-w1601f14\_u\_nobody) (vec00014-eng-cm00003) ACT ... Agent, PAT ... Entity, LOC ... State

keep (EngVallex-ID-ev-w1792f2) (vec00014-eng-cm00005) ACT ... Agent, PAT ... Entity, EFF ... State

keep (EngVallex-ID-ev-w1792f7) (vec00014-eng-cm00006) ACT ... Agent, PAT ... Entity, EFF ... State

keep (EngVallex-ID-ev-w1792f9) (vec00014-eng-cm00007) ACT ... Agent, PAT ... Entity, EFF ... State

hold (EngVallex-ID-ev-w1601f1) (vec00014-eng-cm00012) ACT ... Agent, PAT ... Entity, EFF ... State

keep (EngVallex-ID-ev-w1792f10) (vec00014-eng-cm00015) ACT ... Agent, PAT ... Entity, EFF ... State

Figure 7: Result showing all class members within the given language class

**class ID: vec00382**

**roles: Communicator, Information, Audience, Addressee**

**class: emphasize (ev-w1124f1)**

2 class member(s):

put (EngVallex-ID-ev-w2449f7\_u\_nobody) (vec00382-eng-cm00020) ACT ... Communicator, PAT ... Information, ... Audience, Addressee

```

@id: vec00382-eng-cm00020
@idref: EngVallex-ID-ev-w2449f7_u_nobody
@lang: eng
@status: yes
@lexidref: engvallex
@lemma: put
maparg:
  argpair:
    [0]:
      argfrom: @idref: vecargengACT
              form:
                spec:
    [1]:
      argfrom: @idref: vecroleCommunicator
              form:
                spec:
      argto: @idref: vecargengPAT
            form:
                spec:
      argto: @idref: vecroleInformation
  
```

Figure 8: Result with JSON content expanded

**fine-tune (EngVallex-ID-ev-w1329f1)**

(a) A copy icon for the Sense ID field

Copied EngVallex-ID-ev-w1329f1 to clipboard.

(b) Copied content

Figure 9: Quick copy tool

## 5. Conclusions and Future Work

We have presented a web-based search tool<sup>6</sup> for searching an event-type ontology in general, demonstrated on the linguistically-motivated, richly cross-linked event-type ontology SynSemClass. This tool is aimed at researchers who want to explore the contents of the ontology, make comparisons across the linked external resources or across languages etc. It is linked to a web-based browser that shows all the contents of the ontology in detail, with further options to show or hide contents and to select languages, e.g. for comparison purposes, getting examples and graphical presentation of the results for research papers and other purposes. We will possibly also explore the currently available visualization tools for Linked Open Data, since the ontology could in principle be con-

<sup>6</sup><https://lindat.mff.cuni.cz/services/SynSemClassSearch>

verted to them, but given the specialized nature and structured linking, it might not be possible without some loss of functionality.

The tool will be further developed and generalized to be able to configure it more easily to other similar resources or ontologies. Data will be added as the ontology grows. More complex queries will be designed and implemented, including one going across languages, for example. In the future, once annotated data becomes available, it will be integrated with corpus search.

## 6. Acknowledgements

The work described herein has been supported by the grant *Language Understanding: from Syntax to Discourse* of the Czech Science Foundation (Project No. GX20-16819X). It has been using data and tools provided by the *LINDAT/CLARIAH-CZ Research Infrastructure* (<https://lindat.cz>), supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2023062).

## 7. Bibliographical References

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*, ACL '98, pages 86–90, Stroudsburg, PA, USA. Association for Computational Linguistics.

Claire Bonial, Weston Feely, Jena D Hwang, and Martha Palmer. 2012. Empirically Validating VerbNet Using SemLink. In *Seventh Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, Istanbul, Turkey.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA and London.

Paul Kingsbury and Martha Palmer. 2002. From Treebank to PropBank. In *Proceedings of the LREC*, Canary Islands, Spain.

R. Navigli and Simone Paolo Ponzetto. 2010. Babelnet: Building a very large multilingual semantic network. In *ACL*.

R. Navigli and Simone Paolo Ponzetto. 2012. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.*, 193:217–250.

Kevin Stowe, Jenette Preciado, Kathryn Conger, Susan Windisch Brown, Ghazaleh Kazeminejad, James Gung, and Martha Palmer. 2021. *SemLink 2.0: Chasing lexical resources*. In *Proceedings of the 14th International Conference on Computational Semantics (IWCS)*, pages 222–227, Groningen, The Netherlands (online). Association for Computational Linguistics.

Zdeňka Urešová, Eva Fučíková, Eva Hajičová, and Jan Hajič. 2020. *SynSemClass linked lexicon: Mapping synonymy between languages*. In *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, pages 10–19, Marseille, France. European Language Resources Association.

Zdeňka Urešová, Karolina Zaczynska, Peter Bourgonje, Eva Fučíková, Georg Rehm, and Jan Hajič. 2022. *Making a semantic event-type ontology multilingual*. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1332–1343, Marseille, France. European Language Resources Association.

## 8. Language Resource References

Urešová, Zdeňka and Alcaina, Cristina Fernández and Bourgonje, Peter and Fučíková, Eva and Hajič, Jan and Hajičová, Eva and Rehm, Georg and Rysová, Kateřina and Zaczynska, Karolina. 2023. *SynSemClass 5.0*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, PID: <http://hdl.handle.net/11234/1-5230>.