Speech
Etymology
Idioms
Glossary
NLP
Lemma
Dictionary
Corpora
Meaning
Word
Lexicon
Definition
Headword
Pronunciation
amples
Entry
Lexicology
Dictionary Use
Lexical Resources

EURALEX XIX

Congress of the European Association for Lexicography

Lexicography for inclusion

7-9 September 2021

Virtual

www.euralex2020.gr

2021 Edition

Introducing Terminologue: a cloud-based, open-source terminology management tool

Měchura M. B.[1], Ó Raghallaigh B.[1]

[1] *Fiontar & Scoil na Gaeilge, Dublin City University, Ireland*
valselob@gmail.com, brian.oraghallaigh@dcu.ie

Abstract

This software demonstration introduces Terminologue www.terminologue.org, a cloud-based, open-source terminology management tool. When installed, Terminologue allows users to create, edit and publish termbases via the web. Self-registration is provided, and there are no limits to the number of termbases that can be created or to the number of entries in a termbase. The web-based interface allows registered users to manage their account, to configure their termbases, to modify a termbase's metadata fields, and to edit termbase data. Each entry represents a concept and entries are edited via a tabbed widget which allows users to focus on the different elements of the entry in turn. The overall interface is optimised for both desktop and mobile screens. Data import and export tools are provided, as well as termbase download and upload. Users can be assigned different access levels from read-only to full administrator level. An extranet interface allows lists of entries to be shared with external subject-area experts for review and comment.

Keywords: terminology; terminography; open source; software as a service

1 Introduction

Terminologue is a web-based platform for building, managing and publishing termbases. The software is open-source and can be obtained from GitHub.[1] Documentation is provided for users wishing to install and host their own instance of Terminologue. Alternatively, users who simply wish to use Terminologue can create an account on www.terminologue.org and proceed. This instance of the software is hosted in the cloud by Dublin City University (DCU) and provided as is free of charge. Terminologue is a child of the Léacslann platform (Měchura 2012) and a sibling of the Lexonomy dictionary writing and publishing system (Měchura 2017).

The terminologue software was developed by the Gaois Research Group in Fiontar & Scoil na Gaeilge, DCU, on behalf of Foras na Gaeilge, the state-funded body responsible for the promotion of the Irish language throughout the whole island of Ireland. It was developed to manage the National Terminology Database for Irish (NTD), which is administered by Foras na Gaeilge and contains Irish-language terms approved by the National Terminology Committee (Měchura & Ó Raghallaigh 2010). For this reason, Terminologue is not standards-based but rather was designed to suit the requirements of the NTD, and is based on previous work on this project dating back to its launch in 2005 (Měchura 2006). Nonetheless, it broadly corresponds to established practice, i.e., there are concepts, inside concepts there are terms, definitions, etc. And while Terminologue provides a generic public interface for publishing termbases, the NTD is published via a custom interface available at www.tearma.ie.

Terminologue's back end is written in Node.js, so it runs on both Linux and Windows. Node.js is a JavaScript runtime designed for building scalable network applications.[2] Termbases are stored as SQLite databases and terminological entries are stored internally as JavaScript Object Notation (JSON) data. Terminologue's front end is written in EJS and CSS. The front end uses the Screenful library for screen layout, access management, on-screen interactivity and search.[3]

In addition to Irish and English, the Screenful and Terminologue user interfaces have been translated by users into Czech, Swedish, Welsh, Finnish, Dutch, Russian, Spanish and Arabic. The screen flows from right to left when the Arabic interface is selected.

Terminologue users can create a termbase by clicking on the link to create a termbase on the home page when signed in. Any number of termbases can be created and they can also be deleted. When creating a new termbase, the user gives it a name and a URL. If the user makes their termbase public, it will be available at this URL. Termbases can be created from scratch or from a number of templates. Templates are provided for simple monolingual, bilingual and multilingual termbases, and the settings preconfigured by the templates can be changed later.

Termbases are stored on the server as individual SQLite files. These files can be downloaded via the web interface and users of www.terminologue.org are encouraged to do so to back up their work. Terminologue also allows users to import data from a TBX formatted file, and to export their data into a TBX file, albeit not in a lossless way. Individual entries, lists of entries, or entire termbases can be downloaded in TBX or TXT format. The cloud-based instance of the software currently hosts *c.*750 users and *c.*1,000 termbases. We know from users who have been in touch that it is in use in a number of universities as a teaching tool on translation and terminology courses.

---

[1] https://github.com/gaois/terminologue
[2] https://nodejs.org/en/about/
[3] https://github.com/michmech/screenful

2 Entry Structure

Once a termbase has been created, users can go to its homepage and from there to the editing interface. This is where terminological entries are created and edited. There is a list of entries on the left-hand side. The sample entries can optionally be deleted, and any number of new entries can be created. There is no upper limit on the number of entries a termbase can contain. To open an entry, a user must click on it and it will appear on the space on the right hand-side. What is seen here is a formatted rendering of the entry. This is what the entry will look like if and when it is decided to make the termbase public. To edit the entry, click the Edit button at the top, and Terminologue will open the entry for editing. Regardless of whether or not a termbase is created from a template, Terminologue entries have the same basic structure. This structure comprises administrative data (e.g., whether checked or not), domains, designations (with each designation capable of comprising a term, an acceptability label, a clarification, and sources), intros (i.e., concept disambiguators), definitions, examples, notes, collections, extranets and cross references. Terms comprise an identifier, a language label, text, a list of annotations, and a list of inflections. Each entry represents a concept.
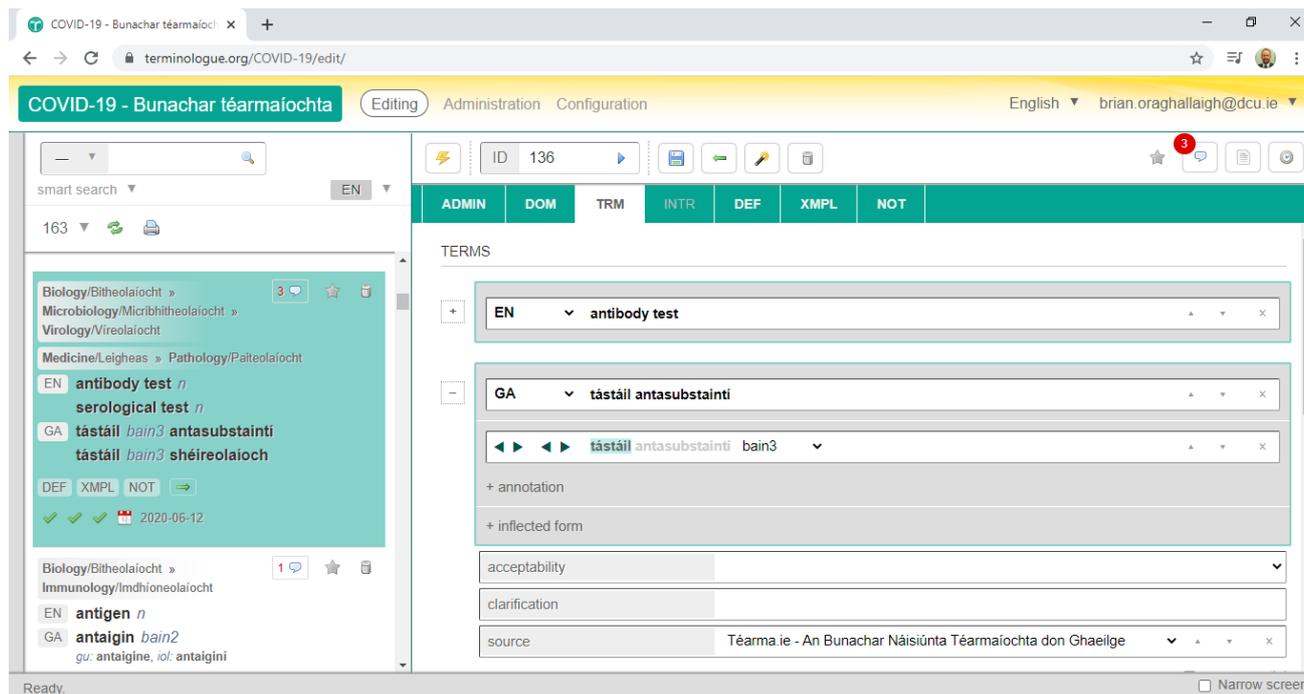

Figure 1: The Terminologue user interface showing the tabbed entry editing widget.

The tabs at the top allow these elements to be edited. The entry's terms can be edited under the TRM tab, its definitions under the DEF tab and so on. For example, to add a new term to the entry, a user must click the plus sign under the TRM tab and fill in the wording of the term. Once the term has been added more information can be added to it, for example a part-of-speech label, or one or more inflected forms. Once a user has finished editing an entry, they have to click the Save button at the top. Entries are stored as JSON data in the *entries* table of the termbase's SQLite database. A copy of the entry JSON is always saved to the *history* table prior to new changes being saved. Older versions of entries can be revived from the history. Each time an entry is saved, indexes stored in the database are automatically updated.

Editing an entry basically means typing text into boxes and selecting values from lists. Many of the lists of values, such as part-of-speech labels, domain labels and so on, constitute the termbase's metadata and can be configured individually for each termbase in the Administration section. If a termbase is started from a template, then it will already have some metadata preconfigured here. Termbase-level settings and functions can be found in the Configuration section. These include user access, name and blurb, languages, alphabetical order, automatic changes (where certain changes trigger automatic administrative labelling of entries), publishing, TBX export and import, SQLite download, URL change, delete and empty.

3 Significant Features

The previous section introduced Terminologue and has, we hope, shown that Terminologue has the features a terminologist typically requires for his or her work. In the next section we are going to zoom in on a few features in Terminologue which we think deserve special attention.

3.1 The Tabbed Editing Interface

When editing a concept, there is a tab for editing its terms, another tab for editing its definitions, and so on: all these items are 'entry-level' items, in terms of the TMF metamodel (Steurs et al. 2015). There is no notion of 'language level' and

'term level' in Terminologue, neither in the user interface nor in the internal data structure. Instead, concepts are composed of items of different types, such as terms and definitions, and these have metadata to indicate what language they are in.

3.2 Optional Sharing of Terms among Concepts

Each individual term can be linked to more than one concept, while any changes made to the term become immediately visible in all concepts it is linked to. This feature is useful because, in the NTD, terms are often polysemous (they refer to more than one concept) and richly grammatically annotated (with part-of-speech labels and lists of inflected forms). Being able to link a term to more than one concept saves work (as it is not necessary to duplicate the grammatical annotations) and prevents inconsistencies.

3.3 Inline Grammatical Annotation

There is a tradition in Irish-language terminography of attaching grammatical labels to individual words inside multi-word terms, for example to the head noun of a noun phrase, as opposed to the entire noun phrase. Terminologue supports this by providing a widget for setting start and end character indexes for annotations. The annotation character indexes are lost when exporting to TBX, however.

3.4 Linguistically Smart Search Features

When making a termbase (or indeed any lexical resource) available to the public online, experience has shown (Měchura 2008) that members of the public often perform searches that are, in one way or another, defective. A high percentage of search requests are inflected forms of words, incorrectly spelled words, or incomplete terms. Linguistically sophisticated algorithms are required to match such search requests successfully to terms in the database. Terminologue has features for spelling error detection (using Levenshtein distance), for lemmatisation (using a large database of inflected forms of words in many different languages), and for partial matching.

3.5 Extranets

Organisations involved in terminological work often enlist the help of external subject-area experts. Cooperation with external experts usually happens on an extranet, which is basically an online environment where the experts can look at (unpublished) terminological entries and add their comments. Terminologue has such a feature, where lists of entries can be viewed and commented on by invited external users, facilitating an integrated workflow from initial drafting to final validation and publishing.

Figure 2: The Terminologue Extranet user interface showing the narrow screen layout.

4 Discussion

Terminologue is the result of a long evolution. The Gaois research group in Dublin City University has been developing terminology management tools since the 2000s, mostly to support the NTD. The NTD could be described as a

public-service terminology database: it is government-funded and its purpose is to disseminate Irish-language terminology among the public. The NTD is also a very popular website, serving over a million searches every month (Měchura & Ó Raghallaigh 2010).

All this means that the NTD, and the software that supports it including Terminologue, has evolved somewhat in separation from the mainstream of terminology work worldwide, which – as it seems to us – is geared mainly towards the needs of corporations and translators. The NTD, on the other hand, targets the general public and its actual user base is wide and broad, encompassing more or less everybody who ever writes non-fiction texts in Irish.

In other words, Terminologue as a terminology management tool is a good fit for public-service terminology work (in any language) but may or may not be a good fit for corporate terminology work, depending on the exact requirements. Some of Terminologue's features exist specifically to support the requirements of public-service terminology work: an example is Terminologue's support for rich grammatical annotation of terms and for the sharing of terms between entries. On the other hand, some features typically found in other terminology software are absent in Terminologue, such as a clear separation between concept-level data, language-level data and term-level data, because a need for this never arises in (our experience of) public-service terminology work.

5   Conclusion and Future Work

In conclusion, we will outline some of the developments we have planned and some possible future directions. Among the developments, closest to production is work to improve the responsiveness and appearance of the interface on smaller screens. Currently users can set the interface to 'narrow screen' mode, but parts of the interface still do not respond well to a narrower screen. We are also looking into streamlining the interface translation process. In addition, users have requested that the public interface include an alphabetic list of all terms in a termbase, to complement the search and term cloud browse functionalities currently available.

With regard to entry structure, we are looking at a number of possibilities. For example, the possibility of including a 'term type' field has been suggested by a number of users and would be compatible with the TMF metamodel and TBX. We are also keen to introduce the option of labelling relationships between concepts. This would allow us to model and visualise conceptual ontologies. The ability to configure default values, to define mandatory fields, and to carry out automatic entry validation have all been mooted. Related to this, we are also looking at ways to better ensure data consistency, to simplify data indexing, and to optimise paging through large termbases.

Terminologue will continue to be actively developed over the next number of years, thanks to financial support from Foras na Gaeilge and institutional support from Dublin City University.

6   References

Měchura, M. B. (2006). Finding the Right Structure for Lexicographical Data: Experiences from a Terminology Project. In E. Corino, C. Marello, C. Onesti (eds.) *Proceedings of the 12th Euralex International Congress*, Torino, 6–9 September 2006. Torino: Edizioni dell'Orso, pp. 1:189–98.

Měchura, M. B. (2008). Giving them what they want: search strategies for electronic dictionaries. In E. Bernal, J. DeCesaris (eds.) *Proceedings of the 13th Euralex International Congress*, Barcelona, 15–19 July 2008. Barcelona: Universitat Pompeu Fabra, pp. 1295–1299.

Měchura, M. B. (2012). Léacslann: A Platform for Building Dictionary Writing Systems. In R. V. Fjeld, J. L. Torjusen (eds.) *Proceedings of the 15th Euralex International Congress*, Oslo, 7–11 August 2012. Oslo: University of Oslo, pp. 855–861.

Měchura, M. B. (2017). Introducing Lexonomy: An Open-Source Dictionary Writing and Publishing System. In *Proceedings of eLex 2017 Conference*, 19-21 September, Leiden, pp. 662–679.

Měchura, M. B. & Ó Raghallaigh, B. (2010). The Focal.ie National Terminology Database for Irish. In *Proceedings of the 14th Euralex International Congress*, 6–10 July 2010. Ljouwert/Leeuwarden, Netherlands.

Steurs, F., De Wachter, K. & De Malsche, E. (2015). Terminology tools. In F. Steurs, H. J. Kockaert (eds.) *Handbook of Terminology: Volume 1*. John Benjamins: Amsterdam, pp. 222–249.