

Distributing Terminology Resources Online: Multiple Outlet and Centralized Outlet Distribution Models in Wales

Gruffudd Prys¹, Tegau Andrews¹, Dewi B. Jones¹, Delyth Prys¹

¹Language Technology Unit, Bangor University, Bangor, Wales, UK
{g.prys,t.andrews,d.b.jones,d.prys}@bangor.ac.uk

Abstract. This paper describes a system which enables the targeted distribution of terms from a single centralized terminology development environment to multiple online outlets. These dissemination outlets include distinct subject-specific websites, a master terminology portal, and aids such as a rollover terminology lookup function that can be added to existing websites.

Keywords: terminology, distribution, dissemination, online, Welsh, dictionary

1 Introduction

A key player in the standardization of terminology in Wales is the Language Technologies Unit (LTU) [1]. Based in the Canolfan Bedwyr centre for Welsh language support at Bangor University, the LTU has been active, in one form or another, since the early 1990s. It marries terminological expertise with technical expertise to enable the creation of online dictionaries and a range of other products, including a Welsh grammar checker, translation memory system, and machine translation system, into which currently over 20 terminology dictionaries can be integrated.

Much of the terminology standardization work currently being undertaken in Wales is accomplished within a centralized online terminology development environment created by the LTU, known as *Maes T*. The most recent online terminology resources in Wales have been generated using this system, including national terminology projects in both Higher Education and in Secondary and Vocational Education [2]. The *Maes T* system is concept-based and language-neutral, and conforms to the ISO standards listed in the guidelines for Welsh terminology work [3]. It was developed in order to facilitate the collaboration of geographically dispersed teams of subject specialists and terminologists, that they might develop, maintain and standardize Welsh-language terms for any and all required domains [4]. *Maes T* is not itself, however, a vehicle for the public dissemination of terms. This paper will discuss an implemented solution to distributing terms from a centralized standardization system to separate, non-homogenous outlets such as websites and web services.

2 Distributing terms

The main commissioners of Welsh-language terminology are public sector organizations which are legally obligated to provide services to the public in both English and Welsh [5]. Such commissioners include the Justice Wales Network, the Local Health Boards and the Welsh Language Board. It is often the case that such clients require

the commissioned terminology resources to be available from their own (possibly pre-existing) websites rather than from a centralized terminology-specific portal such as the Welsh National Terminology Portal [6]¹.

There are a number of reasons for this. Firstly, commissioners may wish to communicate ownership of a terminology resource by having the resource situated on their own web domain, using their own corporate branding. Secondly, the intended user groups for the terminology must be considered. Terminology resources may cater to a number of different audiences, including the professional linguist, the subject specialist, professionals working in a specific domain and the ordinary user of the commissioner's services. Centralized terminology portals that aggregate many different subject fields in a manner that is useful to professional linguists may overload with potentially irrelevant terms those users whose work focuses on a particular domain. As a result, may different user groups may require differing methods of delivering terms.

Thirdly, certain financing bodies require their contribution to specific terminology resources to be acknowledged alongside the resources. On centralized terminology portals which consolidate all terminology resources, this can be difficult to accomplish with sufficient granularity to satisfy the terms of the original grant. Finally, it is important that bodies offering services in a minority language, alongside a dominant language, do so in a prominent manner, giving each language equal visibility on their website. Providing vital terminology on an institution's own websites makes Welsh content easily accessible to the institution's users and can help underline the institution's commitment to operating through the medium of the language.

It is evident therefore that a centralized terminology development system requires the creation of an additional component that can distribute terminology resources to both distinct terminology dissemination outlets and a centralized dissemination outlet such as the Welsh National Terminology Portal. Note that it is not a case of choosing to either disseminate terms through a single central outlet or through many distinct commissioner outlets: most commissioners welcome the chance to make terms available both on their own website and on the Welsh National Terminology Portal.

3 The LTU's distribution solution

The many disparate term distribution outlets in Wales call for an efficient method of distributing both the term data and the language-specific functionalities (such as lemmatization of search queries) which are required to aid users in finding relevant terms.

The solution devised by the LTU was developed using Google Web Toolkit, an open source development toolkit for building and optimizing complex browser-based applications [7]. Google Web Toolkit enables client-side applications to be written in Java and then be deployed as JavaScript. This ensures that the application can be distributed to the external websites of commissioners of terminology work, irrespective of the server technology used, as the it runs within the website visitor's

¹ This portal, named Porth Termau in Welsh, brings together the content of all publicly available terminology dictionaries developed since 1993 by the Centre for the Standardization of Welsh Terminology (now merged with the LTU) and its approved partners.

browser (the vast majority of which are JavaScript compatible) rather than on the website's server.

3.1 Installation on existing websites

Where the commissioner of the terminology work wishes to distribute terms from Maes T on their own pre-existing website, a Javascript "include" can be placed within the header of the web page. This replaces empty elements within the page of the commissioner's website which have been set aside to be the home of the terminology dictionary. However, this requires the co-operation of the website developer, often a subcontractor, who may not always be readily available to the institution.

3.2 Installation on Content Management Systems such as WordPress

When the external website is based on a Content Management System (CMS) such as WordPress, the involvement of the web developer may be avoided by creating a dedicated plugin for the CMS containing the required component code. The LTU uses WordPress, an open source CMS, to create its own term distribution websites in addition to websites that it hosts on behalf of commissioners. WordPress provides a platform that is easy to use and adapt, and it can be localized and adapted for multilingual use using the WPML (WordPress Multilingual) plugin. WordPress' open-source licence and support for plugins enables the platform to be customized as required.

To facilitate the dissemination of terms to WordPress websites, the LTU has created a WordPress plugin, *Porthydd*. The plugin includes code that adds searchable access to the terms found in Maes T to any self-hosted WordPress page that possesses the appropriate API key. The API key, provided by the LTU, establishes whether the website has permission to access the data and controls the settings associated with the search facility and the displaying of search results. These settings include determining which subject-specific terms are displayed, which data fields are displayed, and whether or not to display features such as A-Z lists of a dictionary's contents.

4 An example implementation: *Y Termiadur Addysg*

The example implementation of the LTU's distribution solution chosen to be demonstrated at CHAT 2012 is the *Y Termiadur Addysg* (Education Terminology Dictionary) website. The website is part of a project funded for three years by the Welsh Government with the aim of standardizing and distributing Welsh-language terms to be used in resources such as examinations and course textbooks as well as in general classroom use in primary, secondary and further education in Wales. This is the third project in the *Termiadur* series of terminology dictionaries, and first that will not be published as a print edition.

The *Y Termiadur Addysg* is a fully bilingual Welsh/English online dictionary built on the WordPress platform, featuring guidelines, amended terms, a contact form, a random term feature and game. The search functionality is provided by an instance of the *Porthydd* plugin which can be installed via the WordPress administrator screen by an administrator with very little technical knowledge. *Porthydd* was used to add two pages for accessing terminology to the *Y Termiadur Addysg* website.

The first is a simplified interface that auto-detects the key word language, providing a dual result display for word forms that exist in both languages. The search facility uses language-specific lemmatization rules to return relevant entries even when the search terms feature mutated or conjugated forms. Terms that are alphabetically adjacent to the search term are also displayed in a side panel. The second page features an A-Z letter-by-letter list of all of the dictionary's terms intended to aid those who wish to browse the content of the dictionary. Elements such as abbreviations or parts of speech possess rollover tooltips that provide additional information to the user. Porthydd provides a dynamic link to the terms that are stored in Maes T, so that when a terminologist pushes a button to publish a term in Maes T, it will immediately be available to websites such as Y Termiadur Addysg.

Maes T and Porthydd therefore combine to produce an efficient method of standardizing and distributing terms from a central standardization hub to any number of distribution outlets, including a central multi-domain portal and domain-specific sites.

References

1. Welsh Government: A Living Language: A Language For Living, Welsh Language Strategy 2012-17 (2012), p. 49, <http://wales.gov.uk/docs/dcells/publications/122902wls201217en.pdf>.
2. The Welsh-Medium Higher Education Terminology Project, <http://www.porth.ac.uk/termau>. Y Termiadur Addysg, <http://www.termiaduraddysg.org>.
3. Prys, D., Jones, D. B.: Guidelines for the Standardization of Terminology for the Welsh Assembly Government Translation Service and the Welsh Language Board (2007), <http://www.byig-wlb.org.uk/English/publications/Publications/5338.pdf>.
4. Andrews, T., Prys, G.: The Maes T System and its use in the Welsh-Medium Higher Education Sector in Wales. In: Gornostay, T., Vasiljevs, A. (eds), CHAT 2011: Creation, Harmonization and Application of Terminology Resources, NEALT Proceedings Series, Vol. 12, pp. 49-50. Riga: Northern European Association for Language Technology (2011), <http://dspace.utlib.ee/dspace/bitstream/handle/10062/17370/proceedings.pdf?sequence=1>.
5. Great Britain: Welsh Language Act 1993, Chapter 38. London: Her Majesty's Stationery Office (1993), http://www.opsi.gov.uk/ACTS/acts1993/ukpga_19930038_en_1.
6. Cymdeithas Edward Llwyd, Llên Natur Project, <http://llennatur.com>. Coleg Cymraeg Cenedlaethol (National Welsh College), <http://www.porth.ac.uk>. National Terminology Portal, <http://termau.org>.
7. Google, Google Web Toolkit Overview, <https://developers.google.com/web-toolkit/overview>