

WordNet Website Development And Deployment Using Content Management Approach

*Neha R Prabhugaonkar¹ Apurva S Nagvenkar¹ Venkatesh P Prabhu²
Ramdas N Karmali¹*

(1) GOA UNIVERSITY, Taleigao - Goa

(2) THYWAY CREATIONS, Mapusa - Goa

nehapgaonkar.1920@gmail.com, apurv.nagvenkar@gmail.com,

venkateshprabhu@thywayindia.com, rnk@unigoa.ac.in

ABSTRACT

The WordNets for many official Indian languages are being developed by the members of the IndoWordNet Consortium in India. It was decided that all these WordNets be made open for public use and feedback to further improve their quality and usage. Hence each member of the IndoWordNet Consortium had to develop their own website and deploy their WordNets online. In this paper, the Content Management System (CMS) based approach used to speed up the WordNet website development and deployment activity is presented. The CMS approach is database driven and dynamically creates the websites with minimum input and effort from the website creator. This approach has been successfully used for the deployment of WordNet websites with friendly user interface and all desired functionalities in very short time for many Indian languages.

KEYWORDS: WordNet, Content Management System, CMS, WordNet CMS, IndoWordNet, Application Programming Interface, API, WordNet Templates, IndoWordNet Database, Website.

1 Introduction

WordNet is an important lexical resource for a language which helps in Natural Language Processing (NLP) tasks such as machine translation, information retrieval, word sense disambiguation, multi-lingual dictionary creation etc. WordNet is designed to capture the vocabulary of a language and can be considered as a dictionary cum thesaurus and much more (Miller, 1993), (Miller, 1995), (Fellbaum, 1998). The IndoWordNet is a linked structure of WordNets of major Indian languages from IndoAryan, Dravidian and Sino-Tibetan families. These WordNets have been created by following the expansion approach from Hindi WordNet which was made available free for research in 2006 (Bhattacharyya, 2010).

Most of these language WordNets have reached the necessary critical mass required to open them for public and research use. Feedback from users was the next step to further improve the quality of these resources and increase their usability. Hence, the Consortium decided to make all these WordNets available for public feedback and validation of synsets through online deployment. It was also desirable to have standardisation across all WordNet websites with respect to the user interface, functionality, storage, security, etc. After considering schemes such as Wiki, Blog, Forum and Content Management System we realised that Content Management System was the best option available to publish WordNet content. We also evaluated freely available CMS's like Joomla, Seagull, PHP-Nuke, etc. and concluded that these CMS's were bulky for the task that we set to achieve. From maintenance point of view it was desirable that non-technical person should be able to create and maintain the website with minimal effort and time. So a decision was taken to develop a new CMS for website creation.

The rest of the paper is organised as follows – section 2 introduces the general concept and functionalities of a CMS, section 3 presents the framework and design of our WordNet CMS. The implementation and deployment details of WordNet CMS are presented in section 4. Section 5 presents the results and the conclusions.

2 Content Management System (CMS)

CMS allows you to deploy and manage the content within your website with little technical knowledge.

2.1 Advantages of CMS

The common advantages of CMS are:

- **Designed with non-technical authors in mind:** You can manage the dynamic content of your site with great ease by adding new content, editing existing content or publishing new content.
- **Ease of Maintenance:** In case of traditional websites there is no proper separation of roles when it comes to website developer and content creator. Any changes to the content, menus or links have to be made through an HTML editor. This can at times be difficult for non technical content creator. Absence of “back end” or “admin” feature requires someone with necessary technical skills to make the changes. Without a CMS, the content quality is not properly monitored and becomes inconsistent. CMS gives direct control over the content on the website to the content creator. The back end database stores the content appearing on the website and other important information.
- **Consistent Presentation:** CMS offers well-defined templates for content presentation on the site to maintain a consistent look and feel. Normally, these can be changed and

customised so that the site will have a unique look. It also means that a site redesign can take around ten minutes for the simple changes.

- **Integration of new modules/components:** New features and functionality can be easily integrated as source code for modules which are responsible for all the functionality provided by the website are also maintained in the back end database.
- **Role based permissions:** This helps in content management. Role based access can be provided to create/edit/publish (Author/Editor/Administrator) content.
- **User-friendly interface:** Basic Computer knowledge is required to operate a CMS and therefore no need of specialized technical manpower to manage the website.
- **Control over Meta data in Web page:** You can control the Meta data with appropriate keywords that reflect the content on each page and expected search engine behaviour.
- **Decentralised maintenance:** You do not need specialized software or any specific kind of technological environment to access and update the website. Any browser device connected to the Internet would be sufficient for the job.
- **Automatic adjustments for navigation:** When you create a new navigation item, a new item in a menu, or a new page of any kind, the CMS automatically reconfigures the front end navigation to accommodate it.
- **Security:** The CMS stores information in a database system where access control mechanisms can more easily restrict access to your content. The information is only accessible via the CMS thereby providing better protection for site's content from many common and standard web site attacks.

2.2 Databases used

We have also implemented a relational database to store the WordNet data. This database design (IndoWordNet database) supports storage for multiple WordNets in different languages. The design has been optimised to reduce redundancy. The data common across all languages is stored in a separate database and its size is 1.8 MB. The data specific to a language is stored in the database of respective language. The database size may differ from language to language depending on the synset information. For Konkani the size of this database is 7 MB for thirty thousand synsets. An object-oriented API (IndoWordNet API) has also been implemented to allow access of WordNet data independent of the underlying storage design. The IndoWordNet API allows simultaneous access and updates to single or multiple language WordNets. The heart of the WordNet CMS is a database (CMS database) that stores all the CMS data which is necessary to deploy all the implemented modules. The size of the CMS database is 1 MB for Konkani and should be the same for others.

3 Framework and Design of WordNet CMS

The block diagram of WordNet CMS is shown below in figure 1. An important feature of WordNet CMS is a customizable template, to customize the overall look and layout of a site. A template is used to manipulate the way content is delivered to a web browser. Additionally using CSS within the template design, one can change the colours of backgrounds, text, and links or just about anything that one could within an ordinary XHTML code. The designs for these are all set within the template's CSS file(s) to create a uniform look across entire site, which makes it easy to change the whole look just by altering one or two files rather than every single page (Brinkkemper, 2008).

Template also provides the framework that brings together default functionality and features

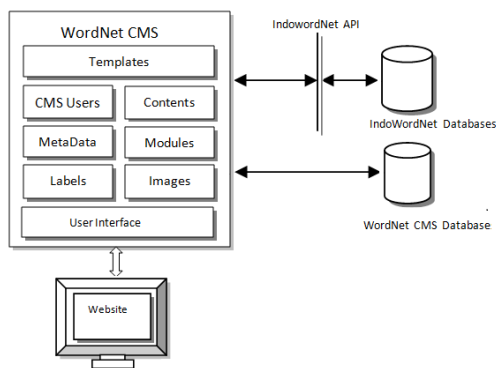


Figure 1: Block diagram of WordNet CMS.

implemented through modules. Functionality and features can be customized or added by the user by customizing the default modules or adding new modules. This offers advantage over traditional websites where such change needs redesign of the entire website. The navigation menus and links are also auto generated to reflect these changes.

3.1 WordNet CMS Modules

A module is an independent component of WordNet CMS which offers a specific functionality. These modules depend on CMS database. While the addition of new modules does not require any changes to the CMS database, new tables may need to be added to store data specific to module functionality. Presently there are six default modules, namely Web Content module, FAQ module, WordNet module, Word Collection module, Terminology module, and Feedback module.

1. **WordNet module:** Provides online access to the WordNet data. The basic functionality supported are search for synsets containing a word, access synsets related through semantic and lexical relationships and compare two or more synsets.
2. **Web content module:** Textual or visual content that is encountered as part of the user experience on websites. A wide range of content can be published using the CMS. This can be characterised as: simple pages, complex pages, with specific layout and presentation and dynamic information sourced from databases, etc. The examples of Web Content are Introduction, About WordNet, About Us, Credits, Contact Us, etc.
3. **Frequently asked questions (FAQ) module:** Listed questions and answers, all supposed to be commonly asked in some context, and pertaining to a particular topic.
4. **Terminology module:** The technical or special terms used in a business, science or special subject domain. For WordNet CMS, it is a vocabulary of technical terms used in Natural Language Processing.
5. **Words Collection module:** The list of all words available in synsets of a particular language. Selecting a word opens its synsets WordNet module.
6. **Feedback module:** Valuable feedback from visitors and users of the website that helps

to improve the overall experience of the site, and its contents. Feedback can range from general visitor's views, comments, and suggestions to discrepancies in synset data and complaints.

The CMS also supports creation of multilingual user interface for the website and customizable on-screen keyboard for all languages. The multilingual user interface is supported through suitably implemented Content and Label components of the CMS. Role based access mechanism is available to restrict access to certain parts and features of the CMS to different users. The WordNet CMS also allows control of Meta data embedded in the generated web page so as to reflect the content on each page as well as provide search engines clues to how the web page should be handled. The WordNet CMS supports both left-to-right and right-to-left text rendition and allows adjustment of the layout as per direction in which content language is written through a simple setting of a flag.

3.2 Architecture of WordNet CMS

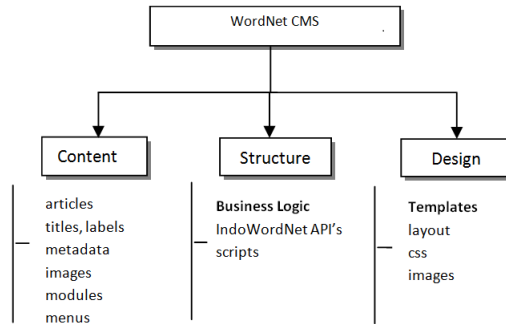


Figure 2: Architecture of WordNet CMS.

As seen in figure 2 above the WordNet CMS is implemented in three layers: Design, Structure, and Content (not to be confused with Content module). The functional division among these three layers allows many advantages throughout the life cycle of the website deployed using CMS. Each layer of the CMS can be recreated and adjusted independent of the other layers. The Design layer can be completely reworked for a new user interface without the need for any adjustments to Structure or Content. The Structure layer can be enhanced for additional functionality with no changes required to Design and Content. Content layer can be changed with no need to adjust the front-end design or functional structure. This three layer architecture makes CMS highly flexible and customizable as per user requirement.

4 Implementation and Deployment Details

The WordNet CMS is developed using PHP scripting language and can be hosted on any Web Server which supports PHP version 5.3.15 and above. Currently MySQL version 5.5.21 is used as database. The CMS development was done using XAMPP on 32 bit Microsoft Windows platform. These softwares can be downloaded from their respective sites. The Konkani WordNet

website created using WordNet CMS has been deployed on Fedora 16 Linux Platform using Apache version 2.2.22 and MySQL version 5.5.21 which come bundled with Fedora 16 Linux Platform.

Conclusion and perspectives

The graph in figure 3 below shows the average number of days required to build and deploy the website using the traditional method and using WordNet CMS. The total time taken to develop WordNet website using traditional method was around 47 days. It took 7 days to design the layout and template of the website, 30 days to implement the website and 10 days for the deployment phase. In case of websites which were deployed using the WordNet CMS, the number of days taken was comparatively very less. For the design phase, it took 2 days to design the layout and template, the structure remains the same and therefore hardly any time was spent on coding and debugging. It took another 2 days for the deployment phase. Therefore the total number of days to create and deploy the website using the WordNet CMS was around 5 days.

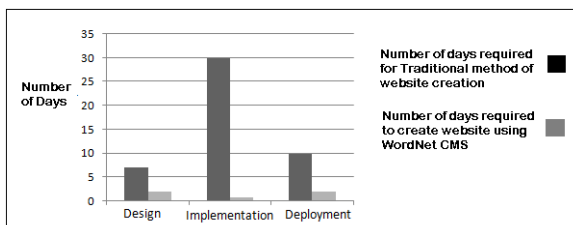


Figure 3: Deployment time requirement analysis.

From the above analysis, we conclude that the WordNet CMS can be used for the speedy deployment of WordNet websites with minimal effort, with good user interface and features by a non technical content creator in a very short time for any language. The enhancements planned for the WordNet CMS are as follows:

1. To develop an installation wizard so that the installation of the CMS is automated.
2. Implementation of Reports module. This will help to keep track of the users visiting the website, provide statistics related to validation of synsets, feedback tracking, etc.
3. Allow further customization of website user interfaces by the website user depending upon the user category such as students, teachers, researchers, linguists, etc. for better user experience.

The WordNet CMS has been successfully used by many IndoWordNet members to design their own language WordNet website.

Acknowledgements

This work has been carried out as a part of the Indradhanush WordNet Project jointly carried out by nine institutions. We wish to express our gratitude to the funding agency Department of Information Technology, Ministry of Communication and Information Technology, Govt. of India and our sincere gratitude to Dr. Jyoti Pawar, Goa University for her guidance.

References

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller (Revised August 1993). *Introduction to WordNet: An On-line Lexical Database*.

George A. Miller 1995. *WordNet: A Lexical Database for English*.

Fellbaum, C. (ed.). 1998. *WordNet: An Electronic Lexical Database*, MIT Press.

Jurriaan Souer, Paul Honders, Johan Versendaal, Sjaak Brinkkemper 2008. *A Framework for Web Content Management System Operations and Maintenance*.

Pushpak Bhattacharyya, Christiane Fellbaum, Piek Vossen 2010. *Principles, Construction and Application of Multilingual WordNets, Proceedings of the 5th Global WordNet Conference (Mumbai-India), 2010*.

