

Web WordNet Interface.

(Un sistema para el desarrollo de WordNets a través de la Web)

Laura Benítez, Sergi Cervell, Gerard Escudero
Mónica López, German Rigau y Mariona Taulé
Grup de Tractament de Llenguatge Natural.
Universitat Politècnica de Catalunya.
Universitat de Barcelona.
{laurab, escudero, g.rigau}@lsi.upc.es
{cervell, lopez, mtaule}@lingua.fil.ub.es

Resumen

Este sistema se ha construido dentro del marco del desarrollo de las wordnets (bases de conocimiento léxicas) del castellano y del catalán¹.

Los datos de estas wordnets están siendo modificados constantemente por los diferentes miembros del grupo al mismo tiempo y además desde diferentes ubicaciones. Esto ha creado la necesidad de garantizar la integridad de estos datos y su consistencia. Para facilitar el trabajo de los usuarios se ha construido una interficie a esta base de datos. Está permite buscar y manipular los datos de manera fácil y efectiva, realizar consultas multilingües que dan una visión intuitiva de las jerarquías y, por último, consultar ciertas fuentes léxicas (diccionarios monolingües y bilingües).

1. Base de Datos

Para minimizar el coste se ha escogido una base de datos de libre distribución.

Por el momento, la base de datos contiene todos los datos de WordNet 1.5., y palabras y glosas asociadas a conceptos para las demás wordnets. Las relaciones semánticas de todas la wordnets son las de la wordnet inglesa. Las wordnets cargadas son fragmentos de las del castellano, vasco y catalán.

Es fácil integrar nuevas wordnets en el sistema y la estructura de la base de datos permite desarrollar software de exportación / importación fácilmente. De hecho, tenemos un programa de exportación e importación de

wordnets monolingües a las herramientas del proyecto EuroWordNet² (Novell, 1996-1998).

2. Interficie

Para construir las wordnets castellana y catalana se ha hecho necesaria la creación de una interficie de acceso a la base de datos y las diferentes fuentes léxicas. Dada la distancia física que separa a los diferentes miembros del proyecto que modifican los datos y para permitir a cualquier usuario consultar los datos abiertamente, se ha desarrollado la interficie para que funcione a través de la Web.

Las operaciones que permite la interficie son las siguientes:

- **Consult Multilingual WordNet.** Su objetivo principal es ayudar en la evaluación de errores y omisiones en las jerarquías de las diferentes wordnets. Para hacer posible esto, la interficie puede mostrar la jerarquía semántica de una de las wordnets, para cualquier relación semántica, cogiendo como punto de partida un concepto, sentido o palabra de una de las lenguas disponibles.
- **Edit Multilingual WordNet.** Su objetivo básico es la edición de las palabras y glosas de todas las wordnet, excepto de la inglesa.
- **Consult Lexical Resources.** Permite consultar los diferentes diccionarios monolingües y bilingües disponibles.
- **Edit Statistics.** Esta opción muestra una lista del histórico de cambios realizados con la interficie.

3. Líneas futuras

Las ampliaciones que se están realizando ahora o que se seguirán en un futuro próximo vienen listadas a continuación:

¹ El sistema presentado en este documento está enmarcado en el contexto de los proyectos ITEM (TIC96-1243-C03-03), EuroWordNet (LE2-4003) y CREL (uno de los centros de referencia de la *Generalitat de Catalunya* para la investigación y el desarrollo).

² Polaris 1.3. de Novell Inc.

- Ampliación de la base de datos y de la interficie para que acepte todos los datos especificados por el proyecto EuroWordNet.
- Adición de la consulta de corpus a la consulta de fuentes léxicas.
- Inclusión de los datos de la clase Levin a los verbos de las wordnets.
- Generador de informes con estadísticas del estado de los datos y de su evolución en el tiempo.

Bibliografía

- Aronson, A. & Lowery, J. (1997). *HTML 3.2. Manual of Style*. - Emeryville: Ziff Davis Press, 1997. - ISBN: 1-56276-529-9.
- Atserias, J.; Climent, S.; Farreras, J.; Rigau, G. & Rodríguez, H. (1997). "Combining Multiple Methods for the Automatic Construction of Multilingual WordNets". In proceedings of Conference on Recent Advances on NLP. (RANLP'97). Tzgov Chark, Bulgaria 1997.
- Benítez, L.; Cervell, S.; Escudero, G.; López, M.; Rigau, G.; Taulé, M. (1998a). "Methods and tools for building the Catalan WordNet". In workshop on Language Resources for European Minority Languages at Conference on Language Resources and Evaluation (LREC). Granada, Spain, 1998.
- Benítez, L.; Escudero, G.; Farreras, J. & Rigau, G. (1998b). "Making Tractable the *Enciclopèdia Catalana* Bilingual Dictionary". Technical Report LSI-98-3-T. LSI Department, *Universitat Politècnica de Catalunya*, 1998.
- Benítez, L.; Escudero, G.; Farreras, J. & Rigau, G. (1998c). "The Application of some Automatic Methods for the Construction of Multilingual WordNets". Technical Report LSI-98-4-T. LSI Department, *Universitat Politècnica de Catalunya*, 1998.
- Benítez, L.; Escudero, G.; Farreras, J. & Rigau, G. (1998d). "WWI: A Multilingual WordNet Interface using the Web. Technical Manual". Technical Report LSI-98-5-T. LSI Department, *Universitat Politècnica de Catalunya*, 1998.
- Benítez, L.; Escudero, G.; Farreras, J. & Rigau, G. (1998e). "WWI: A Multilingual WordNet Interface using the Web. User Manual". Technical Report LSI-98-6-T. LSI Department, *Universitat Politècnica de Catalunya*, 1998.
- CGI. The Common Gateway Interface Specification. [Web Page] "<http://hoohoo.ncsa.uiuc.edu/cgi/interface.html>".
- Cookies. Client Side State - HTTP Cookies. [Web Page] "<http://home.netscape.com/newsref/std/cookiespec.html>".
- Goodman, D. (1996). *Danny Goodman's JavaScript Handbook*. - Foster City: IDG Books Worldwide, 1996. (tr: *Programación en JavaScript*. - Madrid: Anaya Multimedia, 1997. - ISBN: 84-415-0080-0).
- Miller, G. (1990). "Five papers on WordNet". In Special Issue of International Journal of Lexicography: actas 1990.
- Mini SQL. Hughes Technologies. Mini SQL 2.0 User Guide. [Web page] "<http://www.Hughes.com.au/>".
- Muelver, J. (1996). *Creating CoolT M Web Pages with Perl*. - Foster City: IDG Books Worldwide, 1996. (tr: *Creación de páginas Web con PERL*. - Madrid: Anaya Multimedia, 1997. - ISBN: 84-415-0076-2).
- Novell (1996-1998). The EWN Database Tool. Version 1.3.3. Copyright, Novell Inc. 1996-1998.
- Rigau, G.; Atserias, J. & Agirre, E. (1997). "Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation". In proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL'97). Madrid, Spain, 1997.
- Wall, L.; Christiansen, T. & Schwartz, R. L. (1996). *Programming Perl*. - Second Edition. - Sebastopol: O'Reilly & Associates, 1996. - ISBN: 1-56592-149-6.