

WWI 0.0: A Multilingual WordNet Interface using the Web User Manual

L. Benítez, G. Escudero, J. Farreres & G. Rigau

Abstract

In this paper we present a web interface prototype multilingual wordnet. The aim of this interface is to allow the users to construct this multilingual wordnet in the easiest way. The paper describes how to use this interface and how each concrete task can be performed.¹

¹The work reported here is part of the CREL, EuroWordNet (LE4003) and ITEM (TIC96-1243-C03-03) projects.

Contents

Abstract	1
Constents	2
1 Introduction	3
2 Consulting Wordnet	4
3 Editing Wordnet	10
4 Consulting Lexical Resources	14
A Concepts	17
B List of Error Messages	17
References	18

1 Introduction

In recent years the research in the Natural Language Processing (NLP) field has proved the need for extensive and complete Lexical Knowledge Bases (LKBs).

In this way, the English WordNet developed at Princeton University is consolidating as a *standard de facto* for the lexical-semantic representation of the natural language for English. Considering this English WordNet as a reference, new WordNets in some other languages are being built, such as the ones for Galician, Basque, Spanish and Catalan, for instance.

In order to build any monolingual WordNet and make the user work easier, a friendly interface has been created. Firstly, this one allows the user to search and manipulate the data easily and effectively. The user also can consult multilingual data and have an intuitive view of the hierarchy of the synsets. Secondly, some lexical data (mainly monolingual and bilingual dictionaries) are available.

This report explains how to use this interface.

1.1 System's description

The system provides four modes of interaction:

- *Consult* spanish and catalan words combined with nouns and verbs wordnet's synsets.
- *Consult* all the data about nouns and verbs in *wordnet*.
- *Edit* spanish and catalan words combined with nouns and verbs wordnet's synsets.
- *Edit* all the relations between nouns and verbs synsets in *wordnet*. This option is not currently available.

All user actions (as well as data, hour, user) are accounted so all or part of the changes can be cancelled.

Currently, the system cannot edit (create, delete or change) wordnet synsets. Currently, we are using version 1.5 of wordnet.

Although this system can be used from scratch (that is, attaching by hand Spanish and Catalan word to wordnet synsets), a large amount of data has been loaded from bilingual dictionaries (see [Benitez 98a], [Benitez 98b] & [Benitez 98d]).

1.2 Users

To preserve the security of the system, only a list of accounted users can edit the data.

1.3 Logging in

The system is placed in the following URL:

*"http://nipadio.lsi.upc.es/cgi-bin/wwi.pl"*²

If all works, a page like fig. 1 will appear.



Figure 1: Main Page

2 Consulting Wordnet

If the user want to consult, he must to press over the link labeled *Consult Wordnet*. It will appear a page like the fig. 2 shows.

A box can be seen in the screen under the title *Wordnet*. The utility of it is to write in the item that the user is looking for. He can look for *words*, *variants* and *synsets*. He has to say it in the list box placed beside the box. After the options are chosen, the user has to press the button labeled as *lookup*.

The options avaiables are:

²to date several versions of the interface have been implemented: current version is wwi.html

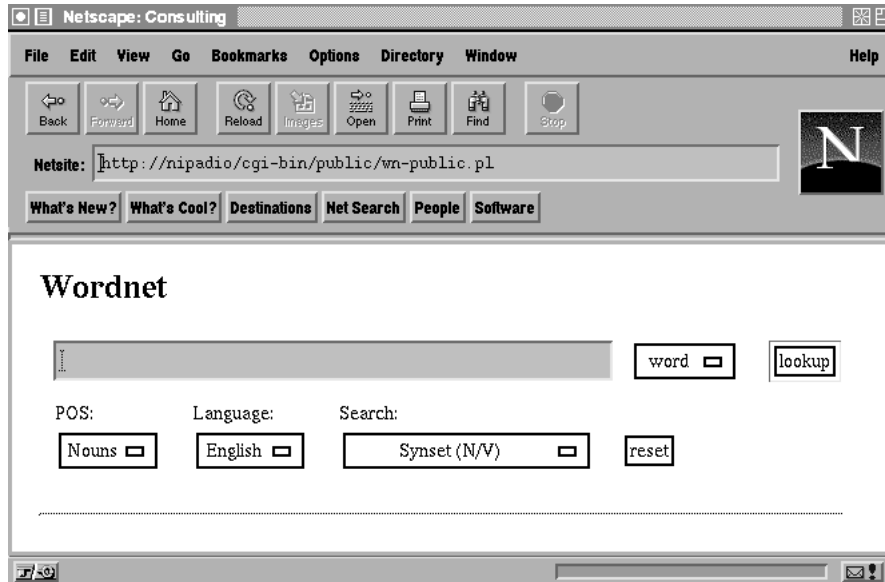


Figure 2: Consult Section

- *POS* (Part Of Speech): it is the POS of the item to search.
 - *nouns*
 - *verbs*
- *Language*: it is the language of the item to search.
 - *english*
 - *spanish*
 - *catalan*
- *Search*: it is the data of the item which the user is asking for.
 - *Synset*: it shows the *synset*, the *semantic file*, the *number of descendants*, all the *english variants* (blue), all the *spanish variants* (red), all the *catalan variants* (green) and the *gloss*. There is an *example* in fig 3.
 - *Hyponyms*: it shows all the *direct hyponyms*. There is an *example* in fig 4. For all the single relations the screen is like the previous one.
 - *Hyperonyms*: it shows all the *direct hyperonyms*.
 - *Part of*: it shows all the synsets which have the item as a part of them. It is a kind of meronymy. This option is only for nouns.

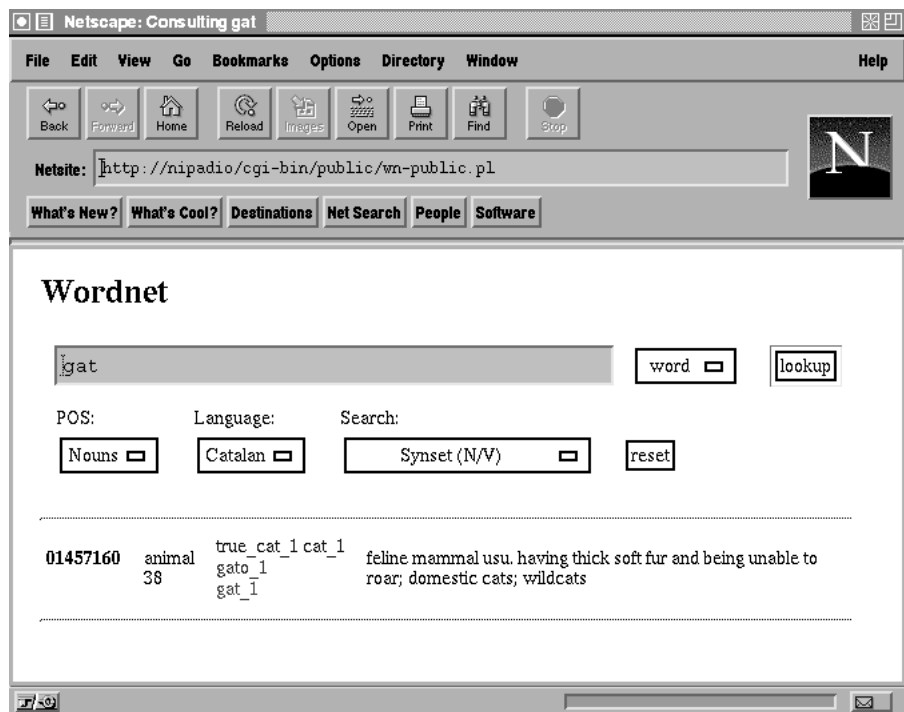


Figure 3: Synset Consult

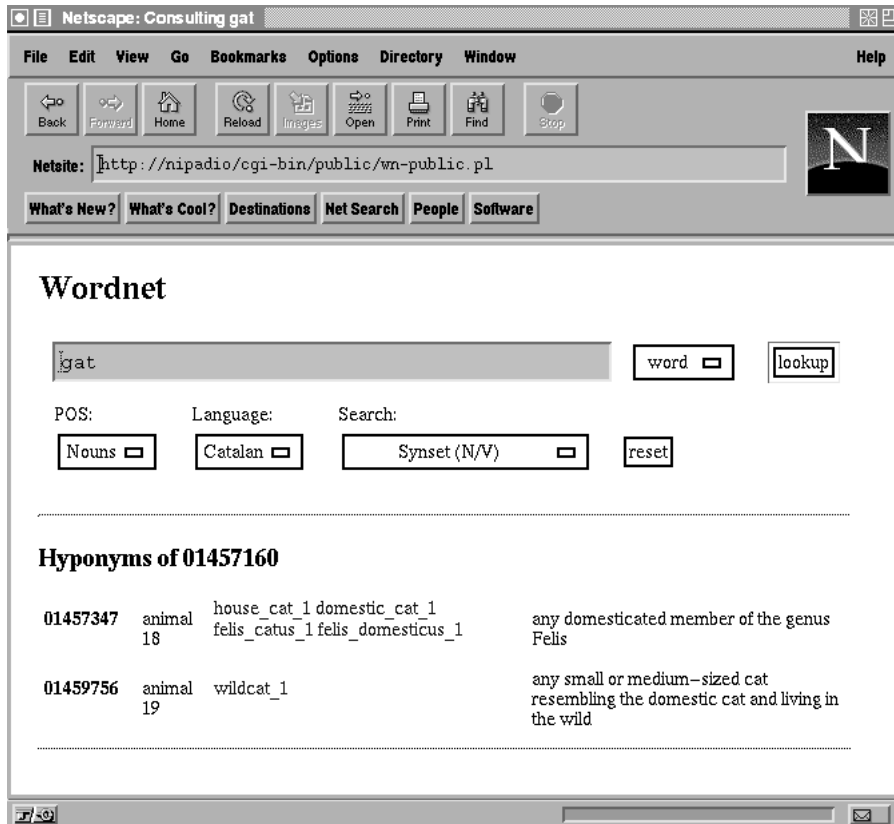


Figure 4: Hyponym Consult

- *Parts*: it shows all the synsets that are parts of the item. It is the inverse relation of the previous one. This option is only for nouns.
- *Member of*: it shows all the synsets which have the item as a member of them. It is a kind of meronymy. This option is only for nouns.
- *Group of*: it shows all the synsets that are members of the item. It is the inverse relation of previous one. This option is only for nouns.
- *Can make*: it shows all the synsets that have the item as a substance of them. It is a kind of meronymy. This option is only for nouns.
- *Made from*: it shows all the substances of the item. It is the inverse relation of the previous one. This option is only for nouns.
- *Antonyms*: it shows the antonyms of the item.
- *Attributes*: it shows the attributes of a noun. This option is only for nouns. There is an *example* in fig. 5.

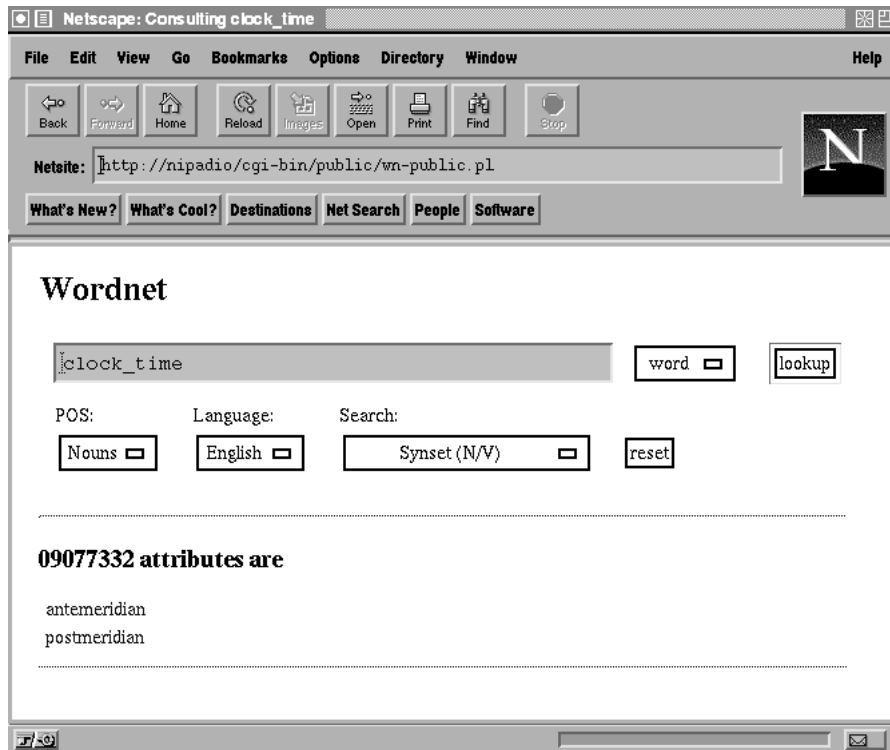


Figure 5: Attributes Consult

- *Entailment*: it shows all the entailed verbs. This option is only for verbs.
- *Cause*: it shows all the caused verbs. This option is only for verbs.
- *See also*: this option is only for verbs.
- *Full hyponyms*: it show the hyponym chain of the item. For an example of screen see *Full hyperonyms*.
- *Full hyperonyms*: it shows the hyperonym chain of the item. There is an *example* in fig. 6.

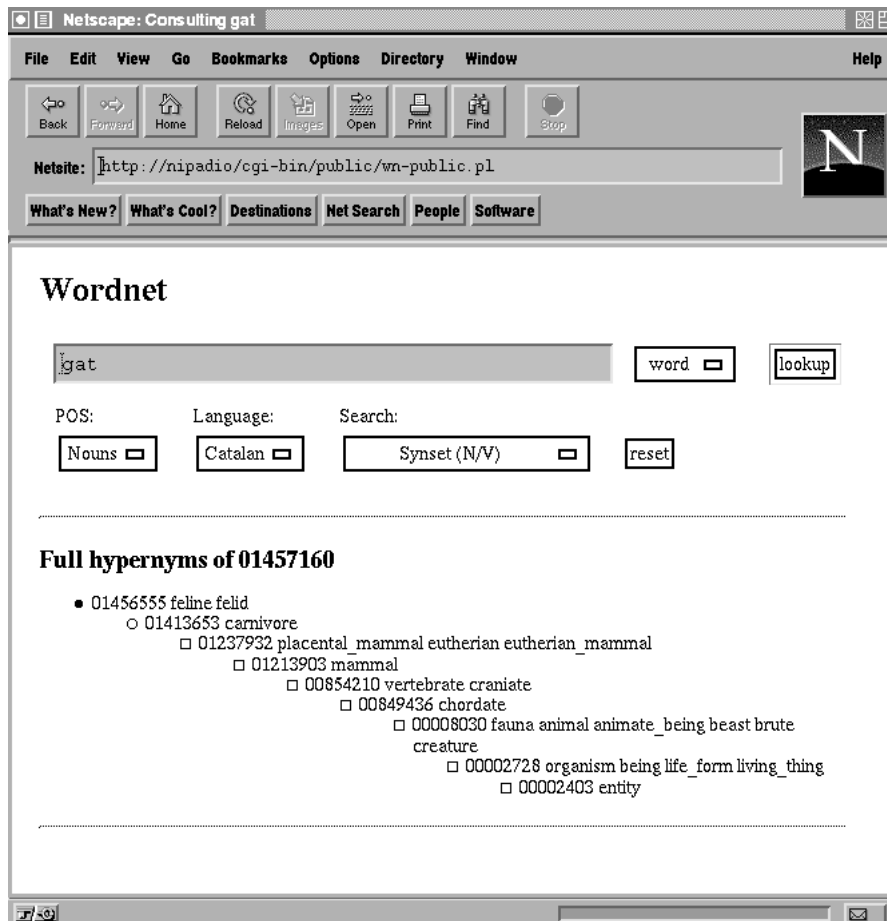


Figure 6: Full-Hyperonyms Consult

- *Spanish hyponyms*: as the full hyponyms option, it shows the hyponym chain of the item, but in this case only those synsets having Spanish words attached are shown.

- *Spanish hyperonyms*: as the full hyperonyms option, it shows the hyperonym chain of the item, but in this case only those synsets having Spanish words attached are shown.

3 Editing Wordnet

For editing the multilingual wordnet, the link labeled as *Edit Wordnet* must be selected. It will appear a dialog box shown in fig. 7.



Figure 7: Dialog box of Edit Section Log In

To continue the process, a *user id* and a *password* are required.

The first page in *edit section* and *consult section* is different in only one thing, the *number of the synset*. For *edit section* appear *buttons* and for *consult section* only *labels*. The fig. 8 shows one of this buttons.

The user can *edit* the *synset* by pressing it. A page like fig. 9 will appear.

In this page the system shows the information of the synset: the *semantic file*, the *number of descendents*, the *english variants* (blue), the *spanish variants* (red), the *catalan variants* and the *gloss*.

Bellow it, it shows a label *Edit:* and a list boxes to select the data to edit. Once an option is selected the user can press the button *Edit*.

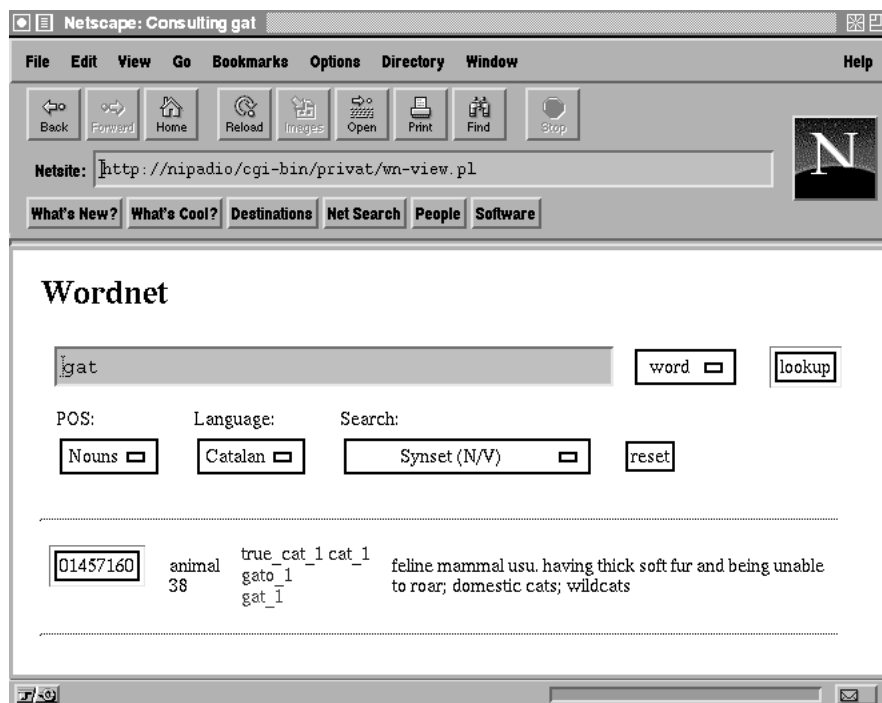


Figure 8: Consult Page in Edit Section

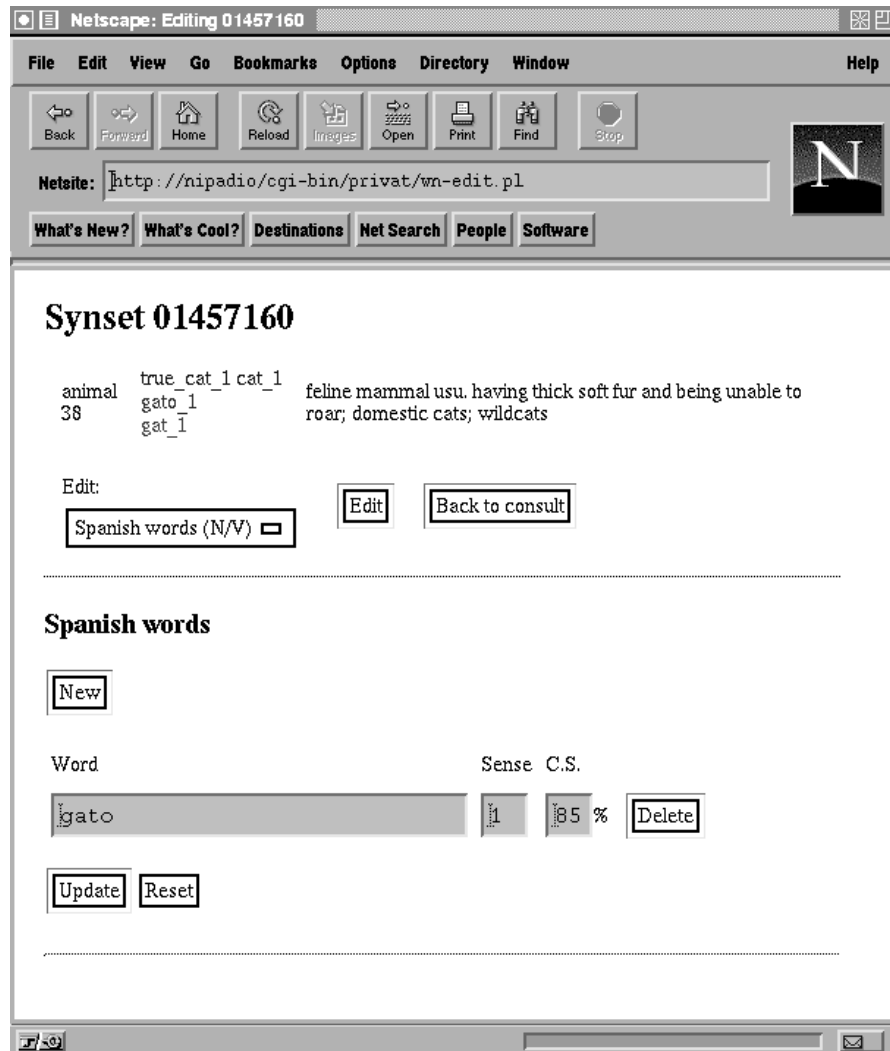


Figure 9: Synset Edit Page

The button *Back to consult* goes to the *consult* page of *edit* section.
Two kinds of information can be edited:

- *Spanish and catalan words*: the screen for both *words* is like fig. 9.
The user can edit the *word*, the *sense* and the *confidence score*.
- *Relations between synsets*: the screen for *relations* is like fig. 10. This part is not currently available.

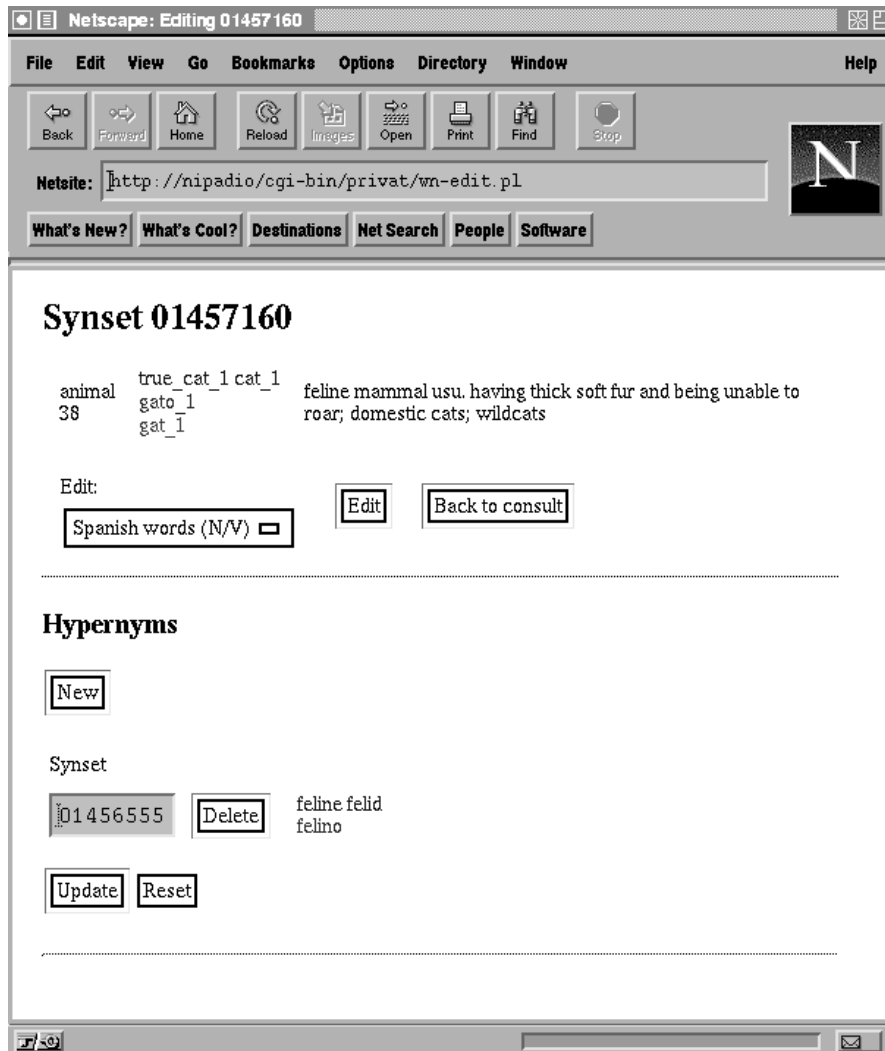


Figure 10: Relation Edit Page

There are in both pages boxes for the data, and for each line a button labeled as *Delete*, it is used to delete the *word* or the *relation*.

The button *New* is used to create boxes for news *words* or *relations*. It is shown in fig. 11.

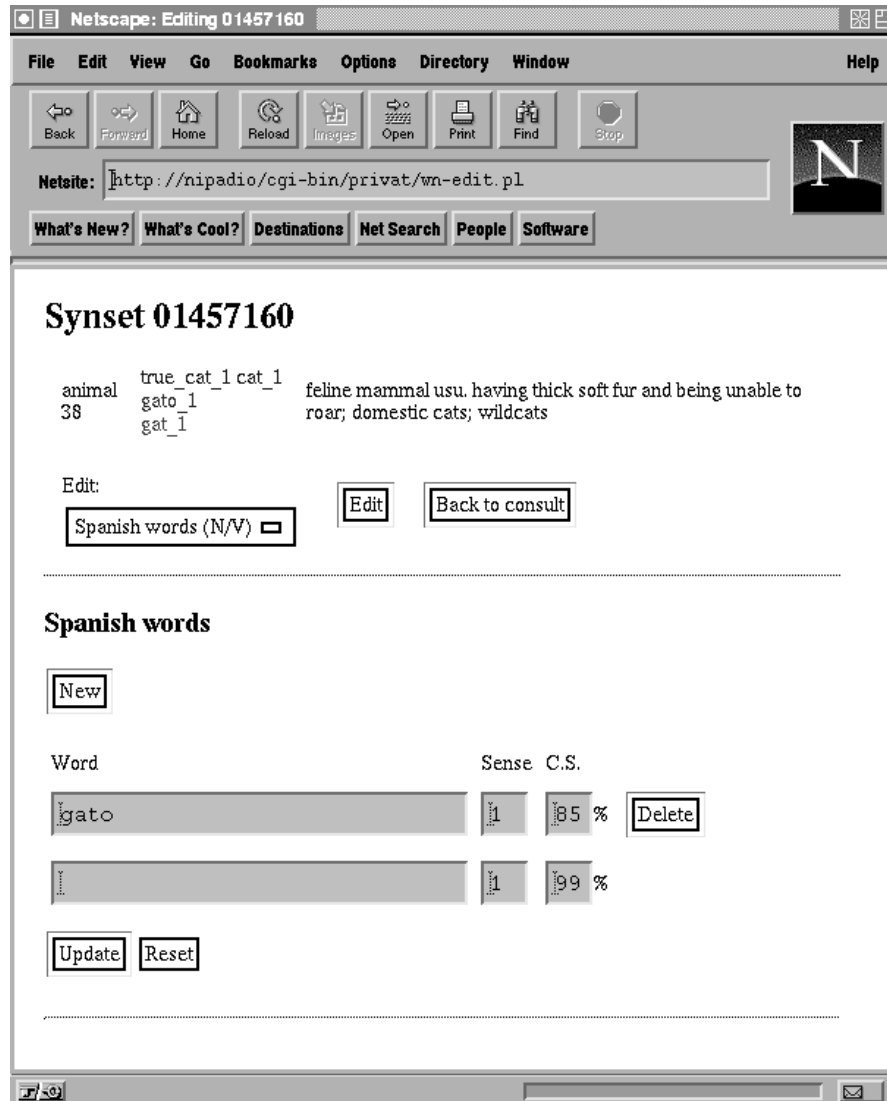


Figure 11: New in Edit Page

The button *Update* is used to update the *changes* and *insert* the data of the new boxes, appearing a screen like fig. 12.

4 Consulting Lexical Resources

If the user wants to consult a lexical resource, he must select the link labeled *Consult Lexical Resources*. I will appear a page like fig. 13.

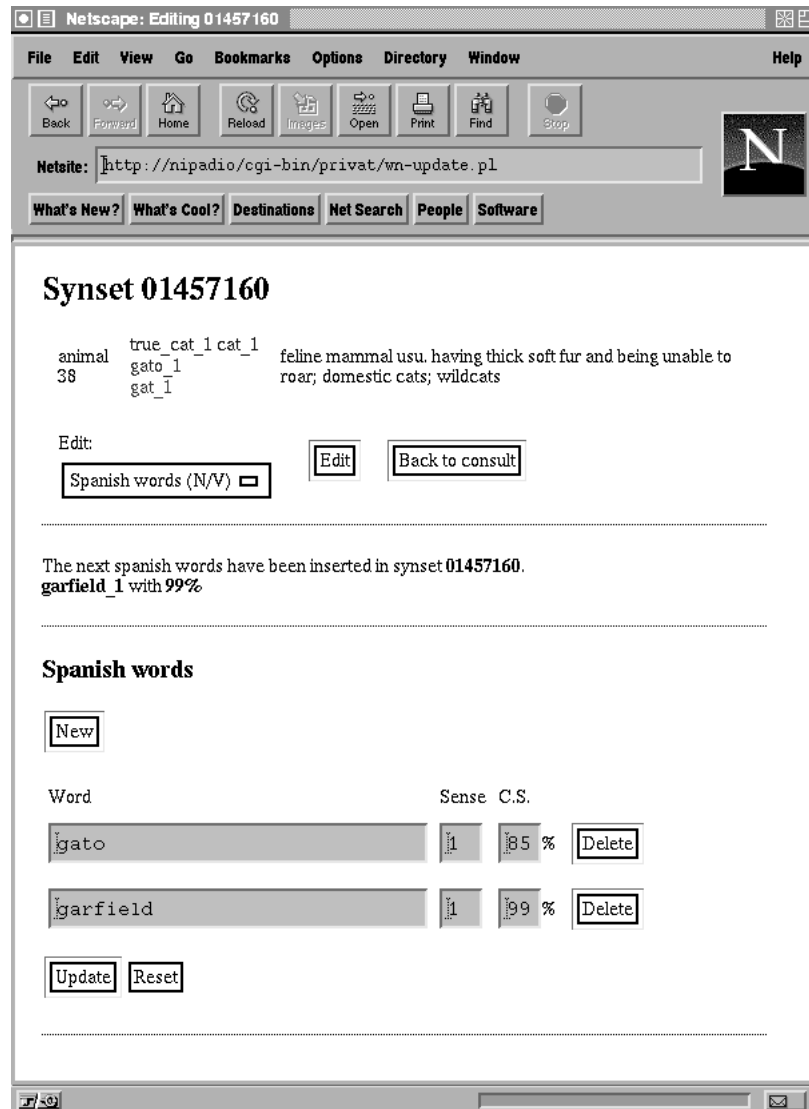


Figure 12: Update in Edit Page

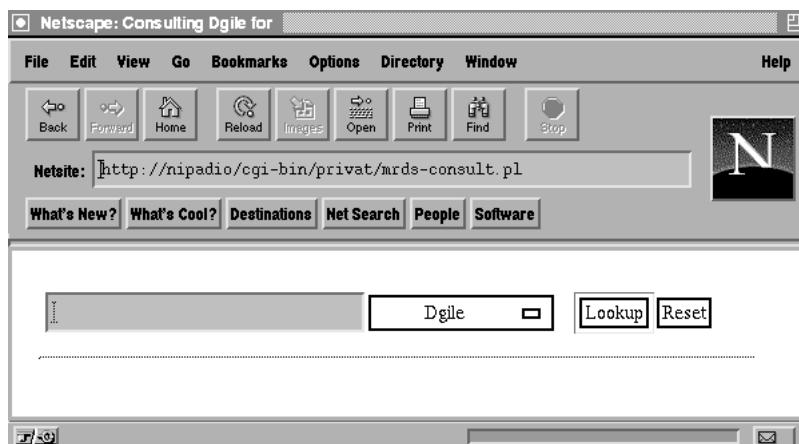


Figure 13: Consult Lexical Resources Section

In this page the user can see a box for writing the word to search. The list box is used to select the lexical resource. The system shows the information as in fig. 14.

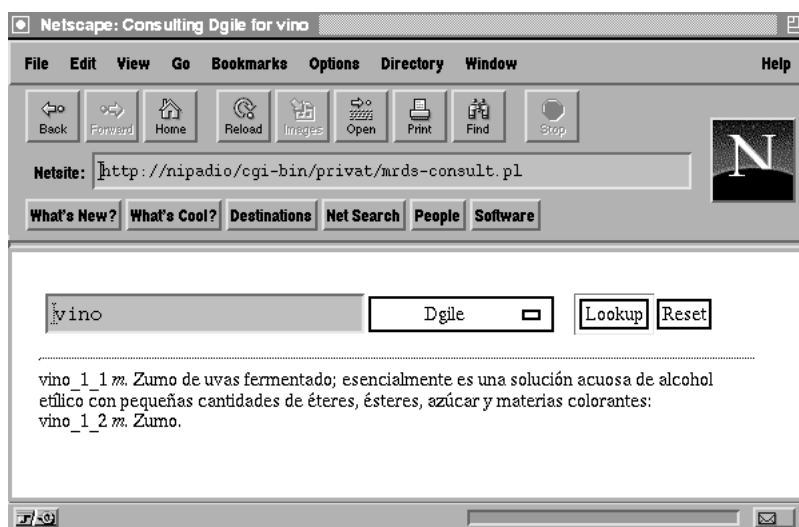


Figure 14: Consult Lexical Resources Section

A Concepts

- *Synset*: is a number that identifies a semantic concept in a POS. Is the acronym of SYNonimy SET. A synset and a word must be unique for a POS [Miller 90].
- *Variant*: is a *word* plus a *sense*. It must be unique in a POS.
- *Sense*: is a number. It enumerate the different semantic values of a word.
- *C.S.*: confidence score.
- *POS*: part of speech.

B List of Error Messages

- *It is only for nouns !!!*: the user has chosen for a verb a search of a noun data.
- *It is only for verbs !!!*: the user has chosen for a noun a search of a verb data.
- *Not found !!!*: the item or data that the user is looking for does not exist.
- *There are no optional fields !!!*: the user has left a box without data.
- *The sense is not correct !!!*: it means that the variant(word_sense) is not unique for his POS.
- *The word XXX already exists for this synset !!!*: it means that the synset_word is not unique for his POS.
- *There is nothing to update !!!*: *updating* no changes has done.

References

- [Aronson 97] ARONSON, Larry; LOWERY, Joseph. *HTML 3.2. Manual of Style*. – Emeryville: Ziff-Davis Press, 1997. – ISBN: 1-56276-529-9.
- [Benitez 98a] Benítez, L.; Escudero, G.; Farreras, J. & Rigau, G. “Applying Automatic Methods for the Construction of Multilingual WordNets”. Technical Report LSI-98-4-T. LSI Department. *Universitat Politècnica de Catalunya*, 1998.
- [Benitez 98b] Benítez, L.; Escudero, G.; Farreras, J. & Rigau, G. “Converting the “Enciclopèdia Catalana” bilingual MTD to an MTD”. Technical Report LSI-98-3-T. LSI Department. *Universitat Politècnica de Catalunya*, 1998.
- [Benitez 98c] Benítez, L.; Escudero, G.; Farreras, J. & Rigau, G. “WWI: A Multilingual WordNet Interface using the Web. Technical Manual”. Technical Report LSI-98-5-T. LSI Department. *Universitat Politècnica de Catalunya*, 1998.
- [Benitez 98d] Benítez, L.; Cervell, S.; Escudero, G.; López, M.; Rigau, G. & Taulé, M. (Universitat Politècnica de Catalunya; Universitat de Barcelona) “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.
- [Goodman 96] GOODMAN, Danny. *Danny Goodman’s JavaScript Handbook*. – Foster City: IDG Books Worldwide, 1996. (tr: *Programación en JavaScript*. – Madrid: Anaya Multimedia, 1997. – ISBN: 84-415-0080-0).
- [Goosens 94] GOOSENS, Michael; MITTELBAACH, Frank; SAMARIN, Alexander *The L^AT_EX Companion*. – [s.l.]: Addison–Wesley, 1994. – ISBN: 0-201-54199-8.
- [Kopka 95] KOPKA, Helmut; DALY, Patrick W. *A Guide to L^AT_EX 2_ε. Document Preparation for Beginners and Advanced Users*. – Second Edition. – Harlow: Addison–Wesley, 1995. – ISBN: 0-201-42777-X.
- [Muelver 96] MUELVER, Jerry. *Creating CoolTM 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).
- [Wall 96] WALL, Larry; CHRISTIANSEN, Tom; SCHWARTZ, Randal L. *Programming Perl*. – Second Edition. – Sebastopol: O’Reilly & Associates, 1996. – ISBN: 1-56592-149-6.

- [Miller 90] MILLER, G. “Five papers on WordNet”. a: Special Issue of International Journal of Lexicography: actas. – [s.l.] [s.n.], 1990.
- [Mini SQL] Hughes Technologies. *Mini SQL 2.0 User Guide*. “URL: <http://www.Hughes.com.au/>”.
- [Cookies] Client Side State - HTTP Cookies. [Web Page]
http://home.netscape.com/newsref/std/cookie_spec.html.
- [CGI] *The Common Gateway Interface Specification*. [Web Page]
<http://hoohoo.ncsa.uiuc.edu/cgi/interface.html>.