

Chooser

Installation Manual

Table of Contents

1. Introduction.....	1
2. Installation.....	2
3. User Interface.....	3
3.1. Textual field.....	4
3.2. List field.....	4
3.3. Information field.....	5

1. Introduction

Chooser is a multi-functional system for corpus annotation. The current implementation have been designed for and employed in word-sense annotation.

The programme maintains different strategies for traversing the text being annotated. The traversal may be carried out over every unit, over non-annotated units, or over all instances of a particular single word or a multi-word expression. Traversal is accomplished through the cursor movement keys (**Arrow keys**) of the keyboard. Using the mouse, the user/annotator is able to mark a particular word/unit in the text and see if and how it has been annotated.

Combining separate graphical units into multi-word expressions is also carried out with the help of the mouse by left-clicking the units in the necessary sequence while pressing the **Ctrl key**. In order to correctly identify the multi-word expression, it is necessary for the MWE lemma to be correctly tagged according to its form in the wordnet database,

which provides the sense inventory for sense annotation.

Chooser can be used for a number of operations concerning corpus editing: wordform editing, lemma editing, deletion or insertion of words in the text if such is necessary (correction of typos, format editing, etc.)

The system provides search options according to word form or lemma. Search may be performed over the whole corpus using Python syntax.

Chooser is coupled with the system for wordnet development, validation and exploration **Hydra** and accesses the senses available in the wordnet database through Hydra's API. The tool's interface embeds a fully-fledged visualisation of the wordnet synonym sets (synsets) – the basic units in a wordnet. The changes made to the wordnet database are dynamically updated and displayed. Any corrections or additions such as newly created synsets and synonyms, textual data, or any other changes, are made accessible to Chooser immediately after they are performed. In such a way the semantic annotation takes place simultaneously with the wordnet development.

2. Installation

To install Chooser, follow the instructions:

(1) Download Chooser from <http://dcl.bas.bg/Tools/Chooser/chooser.zip> and uncompress it into a local directory, e.g. **/home/boby/chooser**.

(2) Chooser depends on the system for wordnet development, validation and exploration **Hydra**. In order to run Chooser, you need to install Hydra first.

Hydra is available for download at: <http://dcl.bas.bg/Tools/Hydra/hydra.zip>. On how to install the programme, create the wordnet database and run the programme, consult Hydra's installation manual:

<http://dcl.bas.bg/Tools/Hydra/Hydra-InstallationManual.pdf>.

(3) Make Hydra available to Chooser.

(a) Put Hydra as a subdirectory 'hydra' of Chooser's main directory, or put it at python's path.

(b) Another option is to make link to it, for example:

```
boby@tornado:~/python/chooser$ ln -s ../hydra
```

(4) Pre-process the corpus and convert it to the required file format described in Section 2.2. of the Chooser User Manual: <http://dcl.bas.bg/Tools/Chooser/Chooser-UserManual.pdf>.

2.2. Starting Chooser

To start Chooser using the command line:

```
boby@tornado:~/python/chooser$ python chooser.py
```

3. User Interface

When the user/annotator loads the file that is to be annotated or edited, Chooser opens a window consisting of three panels:

(1) Textual field where the text that is to be annotated is loaded;

(2) List field with the synsets, from which the annotator chooses the appropriate annotation for each specific unit from the text;

(3) Information field including the different views of Hydra.

3.1. Textual field

The textual field shows the text for annotation. The current unit, which is to be annotated or (in case it is necessary) edited, is marked in a given colour. Different colours mark the characteristics of the units in the text:

- (a) non-traversed units – units in the corpus that have not been traversed in the current session;
- (b) annotated units - language units that are assigned a sense;
- (c) non-annotated units - units that have not been assigned a sense;
- (d) compound words/multiword expressions.

3.2. List field

The list field is a standard list control, which shows the definitions and POS of the synsets among which the annotator chooses in order to annotate the current unit in the text.

Synsets are listed according to the frequency of the previous choices. In this way, the user/annotator sees the most frequently selected synsets in the first rows of the list field. If the user/annotator wants to review the synsets, it can be done either with the help of the keys (by choosing the respective number of the synset, shown at the beginning of the line) or by navigating up and down with the arrow keys and the Enter key or by left-clicking on a definition.

3.3. Information field

The information field includes three views of WordNet, showing different information

about the synset. The synset views correspond to the views of Hydra. This unification of Hydra and Chooser is an important facilitation for the users of both systems and considerably accelerates their work.

For more information about Hydra's synset views see:

<http://dcl.bas.bg/Tools/Hydra/Hydra-UserManual.pdf>

To explore Chooser further, confer the detailed Chooser User Manual:

<http://dcl.bas.bg/Tools/Chooser/Chooser-UserManual.pdf>