

How2Say v1.1  
Installation Manual  
for Terminology Explorer

CONTENT

- 1. Context ..... 2
  - 1.1. Windows/Unix ..... 2
- 2. Installation Procedure ..... 2
- 3. Changing the Document Collections..... 2
  - 3.1. File format of corpus.mflf:..... 2
  - 3.1. How to reload the corpus ..... 3
  - 3.2. Restart and test ..... 3
- 4. Source ..... 3

## 1. Context

---

How2Say is a terminological exploring tool for multilingual corpora. It allows to find for a given expression such as Pommes de terre:

- the super terms « fécule de pomme de terre »
- translation propositions for a target language « potatoes »
- eventually some synonyms « papas (ES) »

How2Say works only with the terms used in the corpora. Consequently, it gives the most exact description of the terminological usage within a specific organization and/or a specific domain. How2Say uses the corpora that can be aligned using the tools of myPREP from the Olanto Suite.

### 1.1. Windows/Unix

---

This installation manual is aimed for Windows and for GNU/Linux (tested on Ubuntu 12.04).

The default installation is done with the “olanto” user. So, the root of the installation will be **/home/olanto/**. After unpacking the archive, you can find in the **config** folder a **forUnix** folder that contains the configuration files suitable for Unix and a **shell** folder that contains the commands suitable for Unix.

The rest of the manual is identical for both systems.

Notations:

For Linux [root] = **/home/olanto; command.sh**

For windows [root] = **C; command.bat**

## 2. Installation Procedure

---

- Prerequisites: both Java8 and TomCat8 must be installed
- Uncompress the **Olanto.zip** file that you have downloaded from Olanto’s website. This file contains a folder: HOW2SAY. The folder must be extracted to the [root].
- Installing the User’s Interface: The following files must be deployed through the Tomcat Manager: o [root]/MYMT/dist/How2Say.war o [root]/MYMT/dist/TermsCorrelation.war
- Starting, Stopping and Re-starting myCAT. After the Tomcat is launched, the HOW2SAY application can be started or re-started by double-clicking this file: Run the **[root]/HOW2SAY/run/ RESTART.sh** or **[root]/ HOW2SAY /run/ RESTART.bat**
- How2Say is provided with a corpus of aligned sentences extracted from English and French Wikipedia articles. (The extraction is performed using myPREP: our tool for aligning comparable corpora)
- Testing: [http://localhost\(:8080\)/How2Say/](http://localhost(:8080)/How2Say/) try this in your browser (adapt the port if need)
- Testing: [http://localhost\(:8080\)/TermsCorrelation/](http://localhost(:8080)/TermsCorrelation/) try this in your browser (adapt the port if need)

## 3. Changing the Document Collections

---

How2Say's corpus of reference documents is located here: C:/MYCAT/corpus/corpus.mflf

You can remove it and replace it with your own corpus.

### 3.1. File format of corpus.mflf:

---

The file corpus.mflf is encoded in URF8.

Each line contains translations of a same sentence for all the languages of the corpus. If a language entry is empty (no translation available), then the content is represented by a dot. A tab separates each translation from the others in the same line. The file is then represented as a table of sentences as follows:

EN	FR	DE	..	ZH
The cat	Le chat	die katze		猫
...	...	...		...

In this example, the language list is “EN, FR, DE, ..., ZH”, the configuration files must be modified to consider this same list of languages.

Modify: **[root]/MYCAT/config/IDX\_fix.xml**

- `<entry key="ROOT_CORPUS_LANG">EN FR DE .. ZH</entry>`

Modify: **[root]/MYCAT/config/IDX\_fix.xml**

- `<entry key="LIST_OF_LANG">EN FR DE .. ZH</entry>`

### 3.1. How to reload the corpus

---

La reconstruction de l’index du corpus se fait en trois étapes (create a new Index, Index the corpus, tag sentence with language). Chaque étape crée un fichier log.

- **Rebuild How2Say:** Run the **[root]/HOW2SAY/run/REBUILD.sh** or **[root]/ HOW2SAY /run/ CreateANewIndex.bat**

### 3.2. Restart and test

---

- Restart the Tomcat,
- After the Tomcat is launched, the HOW2SAY application can be started or re-started by double-clicking this file: Run the **[root]/HOW2SAY/run/ RESTART.sh** or **[root]/ HOW2SAY /run/ RESTART.bat**

## 4. Source

---

The sources are published on Github in myPREP project: <https://github.com/olanto/HOW2SAY>