TreeCloud & Unitex: an increased synergy  
Claude Martineau

To cite this version:  
Claude Martineau. TreeCloud & Unitex: an increased synergy. ECLAVIT Workshop, Nov 2017, Champs-sur-Marne, France. 2017. hal-01702091

HAL Id: hal-01702091  
https://hal-upec-upem.archives-ouvertes.fr/hal-01702091  
Submitted on 19 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
TreeCloud is a tree cloud visualization of a text

TreeCloud builds a tree visualization of a text, which looks like a tag cloud where the tags are displayed around a tree to reflect the co-occurrence distance between the words in the text.

The adjacent grammar contains a path that recognizes auxiliary and modal verbs. The grammar generates a path for this goal. A similar process is used for languages with nouns or adjectives with case (e.g., Serbian or Greek).

2017 on-line version of TreeCloud: an improved implementation

Introduction of the concept of file processing in

The online version of 2014, there is only a single Unix process for each language. The necessary resources were hard-coded into the program.

In the new version, there can be several processing procedures for each language. Furthermore, the language files (Lait and Cyrillic) has been added, in order to manage these two languages (language, language) a concept of processing file has been set up.

For example, in the processing file below, the first line indicates the path for the French resources, then three dictionaries (French general dictionary, first name dictionary, toponym dictionary) have to be applied to the text. In the end, the grammar is applied in the replace mode.

Some examples of trees in different languages

The Tree of the 45 most frequent compound and person nouns in a Jules Verne’s novel

The Tree of the 30 most frequent words in a Serbian article

Conclusion

Plug-in of UniteX into TreeCloud provides:

• A more accurate representation of forbidden words

• A faster construction of the tree (via a careful use of the UniteX API)

http://treecloud.unil-mlv.fr/