

OpenMaTrEx: a free/open-source marker-driven example-based machine translation system

What is OpenMaTrEx?

OpenMaTrEx is a free/open-source (FOS) example-based machine translation (EBMT) system based on the marker hypothesis. It comprises a marker-driven chunker, a collection of chunk aligners, and two engines: one based on the simple proof-of-concept monotone recombinator (previously released as Marclator, <http://www.openmatrex.org/marclator/>) and a Moses-based decoder (<http://www.statmt.org/moses/>).

OpenMaTrEx is a FOS version of the basic components of **MaTrEx**, the data-driven machine translation system designed by the Machine Translation group at the School of Computing of Dublin City University. A great part of the code in **OpenMaTrEx** is written in Java, although there are many important tasks that are performed in a variety of scripting languages.

License

OpenMaTrEx has been released under the GNU General Public Licence (GPL), version 3.

Authors

OpenMaTrEx is © 2007–2010 Dublin City University.

The original **MaTrEx** code was developed among others by Steve Armstrong, Yvette Graham, Nano Gough, Declan Groves, Yanjun Ma, Nicolas Stroppa, John Tinsley, Andy Way, Bart Mellebeek.

The free/open-source package **OpenMaTrEx** has been put together by Sandipan Dandapat, Mikel L. Forcada, Declan Groves, Sergio Penkale, John Tinsley, and Andy Way. Pavel Pecina helped with the Czech marker files.

Funding

M.L. Forcada's sabbatical stay at the Centre for Next Generation Localisation at Dublin City University is supported by Science Foundation Ireland (SFI) through ETS Walton Award 07/W.1/I1802 and by the Universitat d'Alacant (Spain). Support from SFI through grant 07/CE/I1142 is acknowledged.

Downloading

Current versions of the **OpenMaTrEx** package may be downloaded from <http://www.openmatrex.org/>,

where current development may also be followed.

Bug reports, comments, developers welcome

Contact us if you want to send a bug report or a comment, or if you want to become a developer of **OpenMaTrEx**.

Requirements

OpenMaTrEx installs nicely on current GNU/Linux or Unix platforms. A series of free/open-source machine translation packages (GIZA++, Moses, IRSTLM, etc.) have to be installed first. More details in the **INSTALL** file of the package.

Documentation

A more general description of **OpenMaTrEx** may be found in the **ABOUT** file of the package.

Publications

Currently **OpenMaTrEx** implements the baseline components of **MaTrEx** as described in:

- Stroppa, N., D. Groves, A. Way, and K. Sarasola. 2006. Example-based machine translation of the Basque language. In Proceedings of AMTA 2006, pages 232–241. [chunkers, aligners]
- Stroppa, N. and A. Way. 2006. MaTrEx: DCU machine translation system for IWSLT 2006. In Proceedings of the International Workshop on Spoken Language Translation, pages 31–36. [chunkers, aligners]
- Srivastava, A., S. Penkale, D. Groves & J. Tinsley. 2009. Evaluating syntax-driven approaches to phrase extraction for MT. Proc. of the 3rd Intl. Workshop on Example-Based Machine Translation. Dublin, Ireland, pp. 19–28. [translation table merging]

For a more complete list of publications on **MaTrEx**, visit <http://www.openmatrex.org/matrex-publications.html>

Contact

Send your enquiries to:

- Andy Way (away@computing.dcu.ie)
- Mikel Forcada (mforcada@computing.dcu.ie)



Universitat d'Alacant
Universidad de Alicante