Kriya Documentation

Release version

Baskaran Sankaran

October 09, 2014

Contents

Kriya is an implementation of hierarchical phrase-based (hiero) SMT system. See Chiang, 2007

for description of Hiero model.

Kriya is entirely implemented in Python and includes both grammar extractor and decoder modules. Please see the PBML paper for technical details specific to this implementation.

For usage information, see src/README and doc/ directory. The example/ directory includes a toy Fr-En corpus for training along with a Kriya config file for training Hiero grammar. Please read example/README for details first.

If you use this for your research consider citing: Baskaran Sankaran, Majid Razmara and Anoop Sarkar. 2012. Kriya – An end-to-end Hierarchical Phrase-based MT System. The Prague Bulletin of Mathematical Linguistics (PBML), (97), 83–98

Required Software The following is a list of software that might be required depending on what you are trying to do.

- Python - 2.6.2 or later (but not Python 3.0+)
- Giza++ - For word alignments
- MOSES - for extracting phrase alignments from Giza++ output and (optionally) for MERT
- SRILM - For building language model (LM)
- **KenLM - Library for querying the language model (a Python wrapper is included with the decoder** and the source for the wrapper will be released soon)
- MegaM - Mega model optimization package (download from: http://www.cs.utah.edu/~hal/megam/)
- SVMRank - SVM for Ranking optimization (download from: http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html)

At least one of the last two optimization tools are required for Kriya parameter optimization with PRO.

CHAPTER 1

License:

For license information, see the LICENCE file.

CHAPTER 2

Contributors:

Baskaran Sankaran, Majid Razmara and Anoop Sarkar {baskaran,mra44,anoop}@cs.sfu.ca
NatLang Lab, School of Computing Science Simon Fraser University, Burnaby, BC V5A 1S6. Canada

Contents:

CHAPTER 3

Indices and tables

- *genindex*
- *modindex*
- *search*