# Processing text files with ILSP-NLP services and RUTA rules

## Requirements

* Install [Java](https://www.java.com/en/download/)
* Download [this Java client](http://nlp.ilsp.gr/nlp/ilsp-nlp-asclient.jar) and save it, for example, in /home/user/bin/
* Install [Eclipse](https://www.eclipse.org/downloads/) (needed for the RUTA processing step only)
* Install UIMA and UIMA RUTA using these [guidelines](https://uima.apache.org/d/ruta-current/tools.ruta.book.html#section.ugr.tools.ruta.workbench.install) (needed for the RUTA step only)

## Processing text files with ILSP-NLP services

For initial tests, you can download a set of sample txt files from [here](http://nlp.ilsp.gr/nlp/ilsp-nlp-asclient-test-files.zip) or use your own.

Open a shell or a command prompt window and change to a directory with \*.txt files: e.g.

cd /home/user/tmp/txt # or cd c:\tmp\txt for windows.

To process files using the basic processing pipeline [1, 2, 3] (segmentation/tagging/lemmatization/chunking) run:

java -jar /home/user/bin/ilsp-nlp-asclient.jar -b tcp://elrc1.ilsp.gr:61616 -e ilsp-chunker-aggregate-queue -s .txt -a .chunks -ot chunktipster -id ../txt/

The output will be a set of tab separated files that you can examine in any text editor. The last column of these files contains a tag combining the part of speech and several morphosyntactic features for each word. You can read more about the tagset [here](http://nlp.ilsp.gr/nlp/tagset_examples/tagset_en/).

To process files using the basic processing pipeline [1, 2, 3] (segmentation/tagging/lemmatization/chunking) and generate input for the RUTA processing step below, run:

java -jar /home/user/bin/ilsp-nlp-asclient.jar -b tcp://elrc1.ilsp.gr:61616 -e ilsp-chunker-aggregate-queue -s .txt -ot xmicas -a .xmi -id ../txt/

## Processing with RUTA rules

To be added

## References

1. Prokopidis, P., Georgantopoulos, B. & Papageorgiou, H. (2011). [A suite of NLP tools for Greek.](http://nlp.ilsp.gr/nlp/ICGL2011_Prokopidis_etal.pdf) In The 10th International Conference of Greek Linguistics. Komotini, Greece.
2. Papageorgiou, H., Prokopidis, P., Giouli, V. & Piperidis, S. (2000). [A Unified POS Tagging Architecture and its Application to Greek.](http://www.ilsp.gr/administrator/components/com_jresearch/files/publications/LREC_2000_tagger_paper.pdf) In Proceedings of the 2nd Language Resources and Evaluation Conference, pages 1455-1462. Athens. You can find a description of the tagset used by the ILSP FBT POS tagger at [this page](http://nlp.ilsp.gr/nlp/tagset_examples/tagset_en/).
3. Boutsis, S., Prokopidis, P., Giouli, V. & Piperidis, S. (2000). [A Robust Parser for Unrestricted Greek Text](http://www.lrec-conf.org/proceedings/lrec2000/pdf/174.pdf). In Proceedings of the 2nd Language Resources and Evaluation Conference, pages 467-474. Athens.