UIMA/U-Compare Stanford Parser

1. BASIC INFORMATION

Tool name

UIMA/U-Compare Stanford Parser

Overview and purpose of the tool

Syntactic parser for English. Outputs dependency relations. Also outputs parts-of-speech for each token.

The tool is provided as a UIMA[1] (Ferrucci et al., 2006) component, specifically as Java archive (jar) file, which can be incorporated within any UIMA workflow. However, it is particularly designed use in the U-Compare text mining platform (Kano et al., 2009; Kano et al., 2011; see separate META-SHARE record), since the types of annotations it produces are compliant with U-Compare type system.

A short description of the algorithm

The Stanford parser[2] (Klein & Manning, 2003a, Klein and Manning 2003b; de Marneffe et al. 2006) is a Java implementation of probabilistic natural language parsers, both highly optimized PCFG and lexicalized dependency parsers, and a lexicalized PCFG parser. The lexicalized probabilistic parser implements a factored product model, with separate PCFG phrase structure and lexical dependency experts, whose preferences are combined by efficient exact inference, using an A* algorithm. The output is in the form of Stanford dependencies[3]

2. TECHNICAL INFORMATION

Software dependencies and system requirements

The tool is provided as a UIMA component wrapped around a web service. Thus, the tool must be run within the Apache UIMA framework. Alternatively, it can be run within the U-Compare framework. The component has been specifically designed to work in U-Compare workflows and is compliant with the U-Compare type system.

Installation

The tool is provided as an in-built component of the U-Compare workbench. However, it can also be used in other UIMA workflows. Since it is packaged as a UIMA component, no specific installation is required, following installation of the UIMA framework and/or U-Compare.

[1] http://uima.apache.org/

[2] http://nlp.stanford.edu/software/lex-parser.shtml

[3] http://nlp.stanford.edu/software/stanford-dependencies.shtml

Execution instructions

The tool can be used within U-Compare simply be dragging and dropping it into a workflow using the graphical user interface of the U-Compare workbench. Alternatively, it can be incorporated into other UIMA-based workflows, by following the documentation on the Apache UIMA site. Given that the UIMA component is implemented in Java, the tool is platform-independent.

Input/Output data formats

Input data formats

The input is plain text document that has previously been read into the UIMA Common Analysis Structure (CAS) via a UIMA collection reader component. As a prerequisite, the CAS must contain sentence annotations. Thus, a sentence splitter must be executed in the workflow prior the execution of the Stanford Parser component.

Output data format

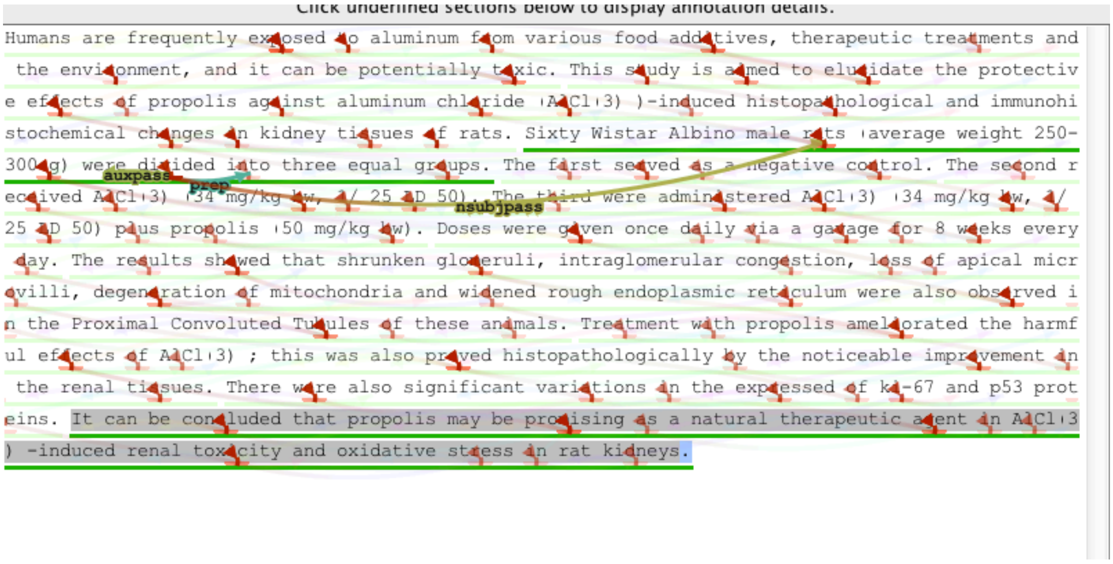
The tool adds POSToken annotations (to encode tokens with their parts-of-speech) and StanfordDependency annotations (to encode dependency relations) to the CAS.

Integration with external tools

As mentioned above, the tool can only be run within the UIMA or U-Compare frameworks.

3. CONTENT INFORMATION

Figure 1 shows the output of the tool in the U-Compare workbench. The sample text is taken the PubMed website (http://www.ncbi.nlm.nih.gov/pubmed/23172825l). The arrows between the words show the dependency relations involving the word "divided".



Click underlined sections below to display annotation details.

```
Humans are frequently exposed to aluminum from various food additives, therapeutic treatments and
 the environment, and it can be potentially toxic. This study is aimed to elucidate the protectiv
e effects of propolis against aluminum chloride (AlCl(3))-induced histopathological and immunohi
stochemical changes in kidney tissues of rats. Sixty Wistar Albino male rats (average weight 250-
300 g) were divided into three equal groups. The first served as a negative control. The second r
eceived AlCl(3) (34 mg/kg bw, 1/ 25 LD 50). The third were administered AlCl(3) (34 mg/kg bw, 1/
25 LD 50) plus propolis (50 mg/kg bw). Doses were given once daily via a gavage for 8 weeks every
 day. The results showed that shrunken glomeruli, intraglomerular congestion, loss of apical micr
ovilli, degeneration of mitochondria and widened rough endoplasmic reticulum were also observed i
n the Proximal Convoluted Tubules of these animals. Treatment with propolis ameliorated the harmf
ul effects of AlCl(3) ; this was also proved histopathologically by the noticeable improvement in
 the renal tissues. There were also significant variations in the expressed of ki-67 and p53 prot
eins. It can be concluded that propolis may be promising as a natural therapeutic agent in AlCl(3
) -induced renal toxicity and oxidative stress in rat kidneys.
```

Figure 1: Output of the Stanford in the U-Compare workbench

Running the tool on the 1 KB text on a single core machine with 8 GB RAM takes around 3254 milliseconds.

3. LICENCES

a) The Stanford Parser UIMA wrapper is licensed using the GNU General Public License version 2.0 (GPLv2). Please see “LICENCE.txt” in the “Licences” directory. Please acknowledge the National Centre for Text Mining, University of Manchester if you use the Stanford Parser UIMA component

b) The underlying Stanford Parser software is licensed using the GNU General Public License version 2.0 (GPLv2). Please see “LICENCE.txt” in the “Licences” directory.

c) The UIMA framework is licenced using the Apache licence. Please see “Apache-licence.txt” in the licenses directory.

4. ADMINISTRATIVE INFORMATION

Contact

For further information, please contact Sophia Ananiadou:
sophia.ananiadou@manchester.ac.uk

5. REFERENCES

Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430

Dan Klein and Christopher D. Manning. 2003. Fast Exact Inference with a Factored Model for Natural Language Parsing. In *Advances in Neural Information Processing Systems 15 (NIPS 2002)*, Cambridge, MA: MIT Press, pp. 3-10.

Marie-Catherine de Marneffe, Bill MacCartney and Christopher D. Manning. 2006. Generating Typed Dependency Parses from Phrase Structure Parses. In *LREC 2006*.