

stav: text annotation visualiser

Pontus Stenetorp*, Goran Topić*, Sampo Pyysalo†‡, Sophia Ananiadou†‡ and Jun'ichi Tsujii†‡§

*Department of Computer Science, University of Tokyo | †School of Computer Science, ‡National Centre for Text Mining, University of Manchester | §Microsoft Research Asia
 {pontus, goran, smp}@is.s.u-tokyo.ac.jp, sophia.ananiadou@manchester.ac.uk, jtsujii@microsoft.com

Visualisation

- Insights into annotations are relevant for:
 - Constructing Natural Language Processing (NLP) systems
 - Curating NLP resources
- Visualisations should be:
 - Intuitive, at-a-glance
 - Accessible, to be discussed with colleagues

First rule: **“Look at the data”**

(Although the second rule is: “Don't look at the data”)

– Freely after Eugene Charniak

stav text annotation tool (stav)

- Initially introduced to visualise biomolecular annotations in the BioNLP 2011 Shared Task (Stenetorp et al. 2011)
- Well-received by the shared task participants
- Generalised to support a wide range of annotation tasks:
 - Web-based
 - Error-analysis capabilities
 - End-user facing visualisation
 - Format conversion tools
 - Full Unicode support
 - File-based back-end, GREP/SED-able
 - Each annotation reachable by a URI

Annotation Primitives

- stav uses the following primitives to represent annotations:
 - Typed spans, as applied in e.g. Named Entity Recognition
 - Binary relations, as applied in relation extraction
 - N-ary associations of annotations, such as event structures
 - Binary or multi-valued attributes on any annotation, e.g. NEGATION or CONFIDENCE
- stav is distributed with several conversion tools for popular formats such as CoNLL-X and BIO

Acknowledgements

This work was supported by the Swedish Royal Academy of Sciences and by Grant-in-Aid for Specially Promoted Research (MEXT, Japan). The UK National Centre for Text Mining is funded by the UK Joint Information Systems Committee (JISC).

Visualisation Examples

Zo was er het voorstel voor een gezamenlijke lijst met Inzet, waarmee
 de VLD in Dendermonde de meerderheid vormt en die bestaat uit SP'ers,
 VU'ers en enkele onafhankelijken.

Figure: Named Entity Annotations

そして、必要な責任は早急に取る決断が望まれる。

Figure: Verb Frame Annotation

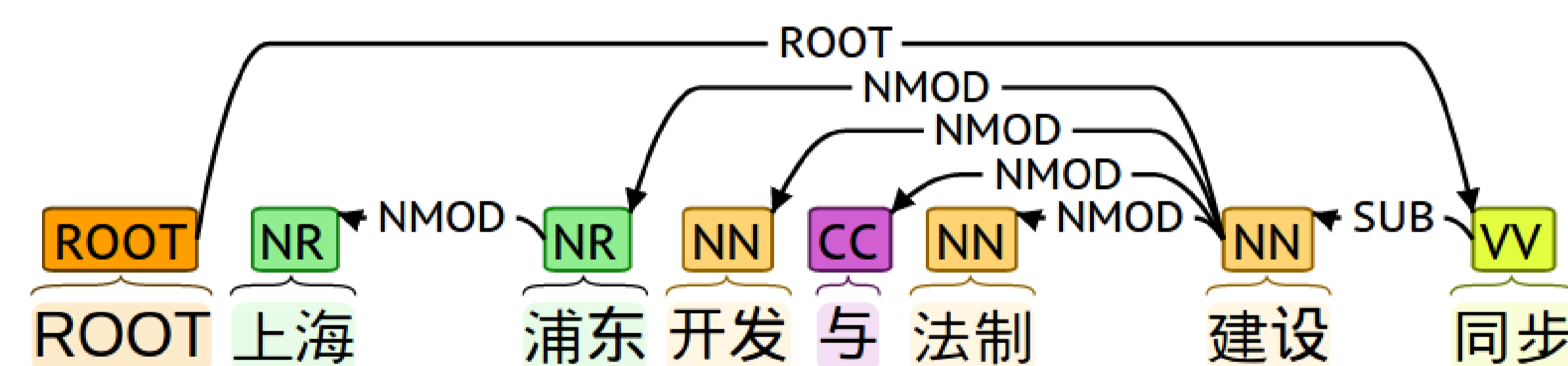


Figure: Dependency Annotation

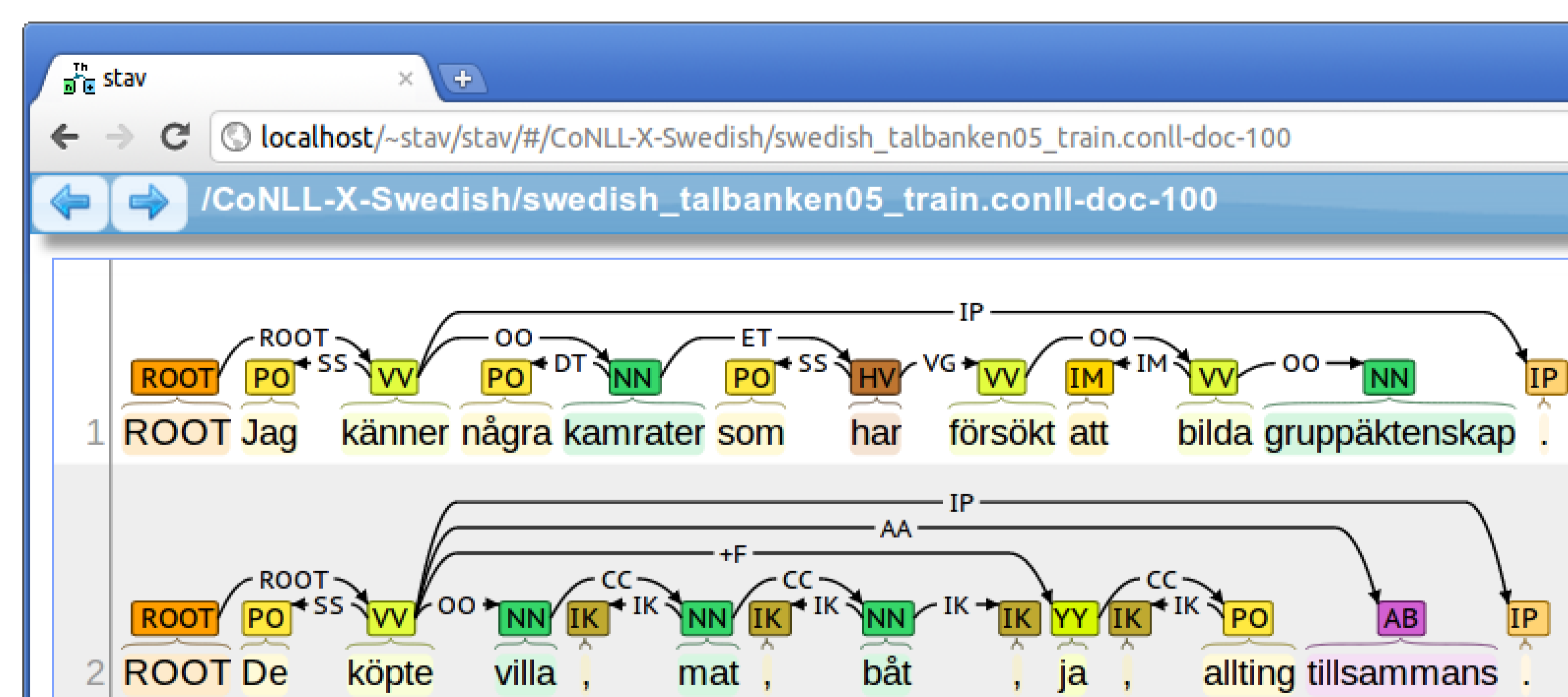
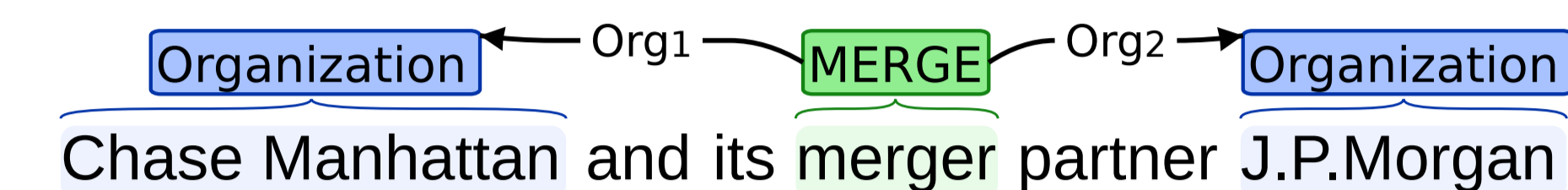


Figure: User-interface

Annotation Format

Visualisation:



Text:

Chase Manhattan and its merger partner J.P.Morgan...

Annotation:

ID	Type	Offsets	Text excerpt
T1	Organization	0 15	Chase Manhattan
T2	Organization	35 49	J.P.Morgan
T3	MERGE-ORG	24 30	merger
E1	MERGE-ORG:T3	Org1:T1 Org2:T2	

Text-bounds Event

Event type and trigger Event Arguments

Search Functionality

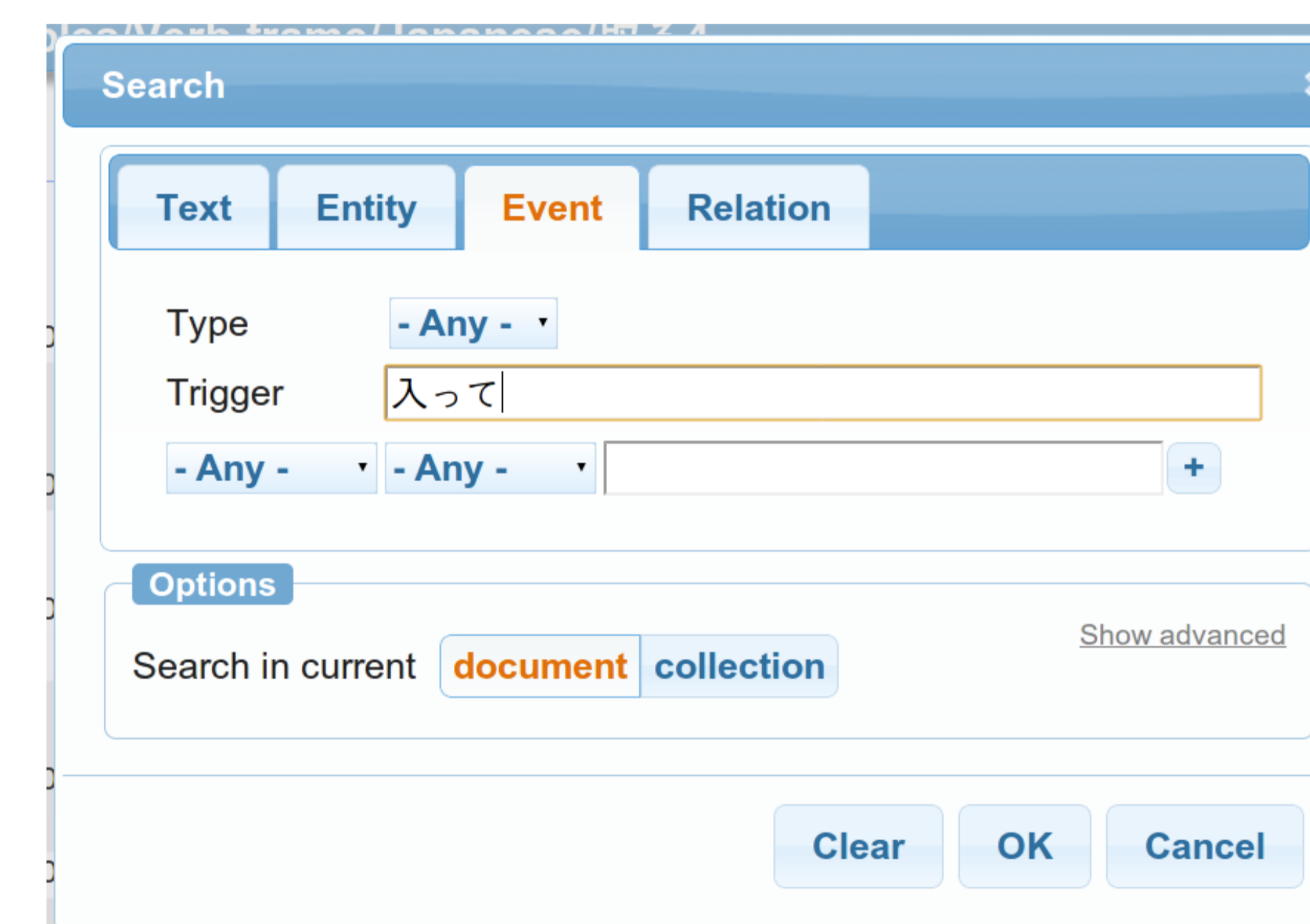


Figure: User-interface for Search Functionality

Availability

Source code, conversion tools and future additions is/will be available at:

<http://github.com/nlplab/stav/>