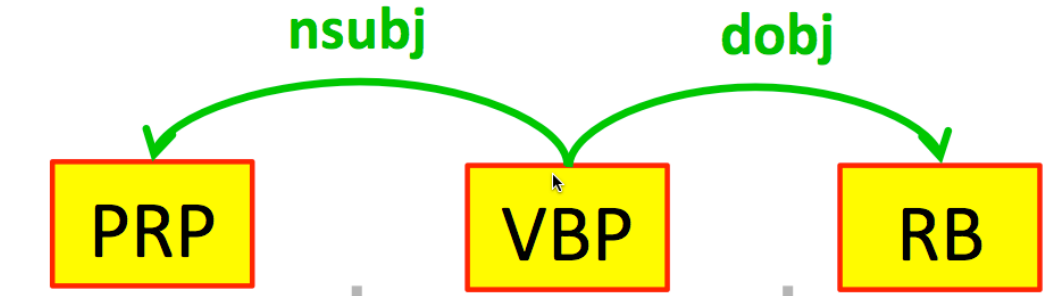
WebAnno: a flexible, web-based annotation tool for CLARIN

Richard Eckart de Castilho, Chris Biemann,
Iryna Gurevych, Seid Muhie Yimam

TECHNISCHE UNIVERSITÄT DARMSTADT



#WebAnno

WebAnno

SPONSORED BY THE
Federal Ministry of Education and Research

CLARIN-D

UBIQUITOUS KNOWLEDGE PROCESSING

Language Technology

UKP

WebAnno – an annotation tool for text

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Team tool**
  - Allows a distributed team of annotators to work on a corpus
  - Supports different roles within the team (e.g. user / manager)
- **Flexible**
  - Multi-layer annotation with configurable annotation layers
  - Different annotation modes including correction and learning modes
- **Web-based**
  - Available to annotators everywhere, no installation effort
  - All configuration performed through the web interface
- **Platform independent**
  - Platform independent Java-based application
- **Open source**
  - Allows the community to participate

WebAnno – an annotation tool for CLARIN

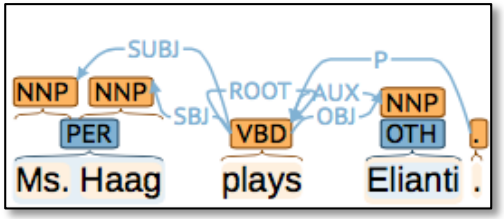
TECHNISCHE UNIVERSITÄT DARMSTADT

- **Developed based on the requirements of CLARIN F-AG 7…**
  - Dipper et al. NoSta-D: A corpus of German non-standard varieties**.** *Non-Standard Data Sources in Corpus-Based Research (2013): 69-76.*
  - Benikova et al. NoSta-D Named Entity Annotation for German: Guidelines and Dataset. *Proceedings of LREC. 2014.*
- **… but also used beyond F-AG 7**
  - Pedersen et al. Semantic Annotation of the Danish CLARIN Reference Corpus. *Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Sem. Annotation. 2014.*
- **… used and recognized beyond CLARIN**
  - Search “WebAnno” on Google Scholar
  - See our public users mailing list
- **WebAnno is the first annotation tool to supporting WebLicht TCF**
  - Worked with TCF developers to improve TCF support updating files!
- **WebAnno team is constantly in touch with the community**
  - Visit http://webanno.googlecode.com after the talk to participate in our survey!

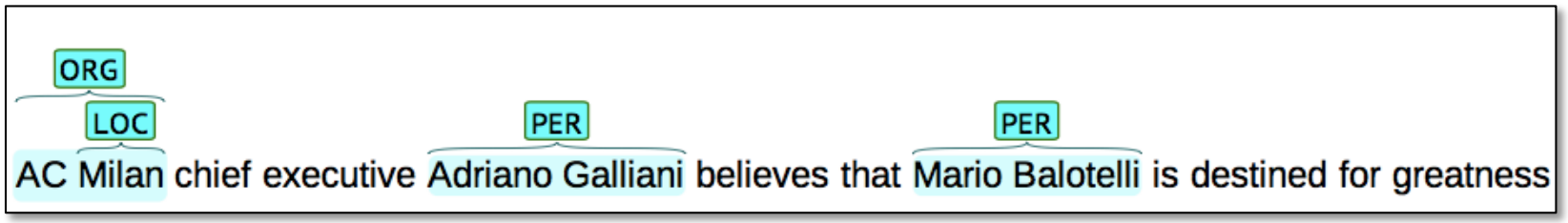
Annotation examples

TECHNISCHE UNIVERSITÄT DARMSTADT

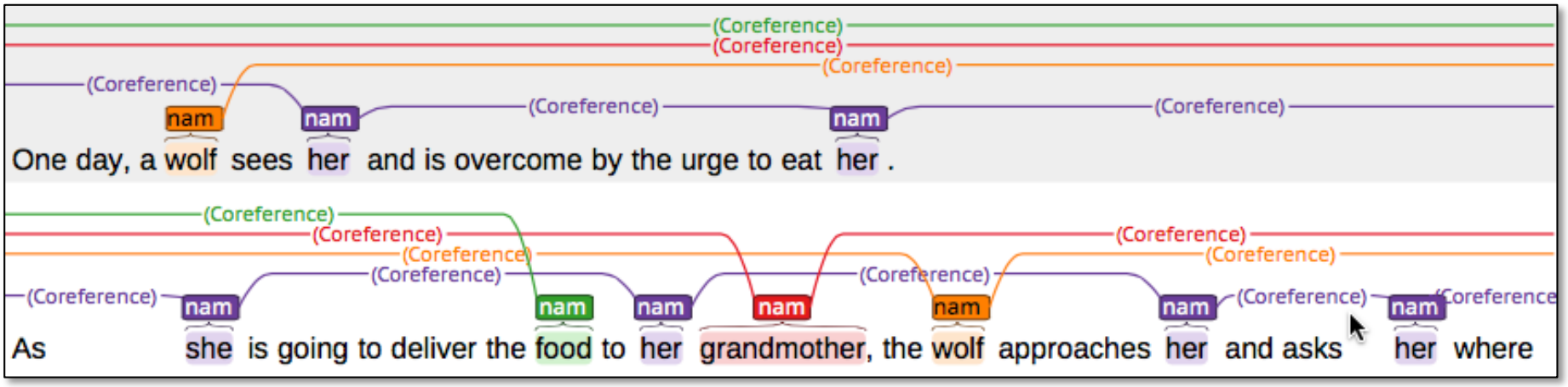
Part-of-Speech & syntactic dependencies



Named entities



Co-reference



Main Menu

TECHNISCHE UNIVERSITÄT DARMSTADT


Annotation

- Annotate texts from scratch


Correction

- Review and correct previously annotated documents


Automation

- Employ integrated machine learning capabilities


Curation

- Compare annotations from different annotators and merge them


Monitoring

- Assign workload to annotators and monitor their progress

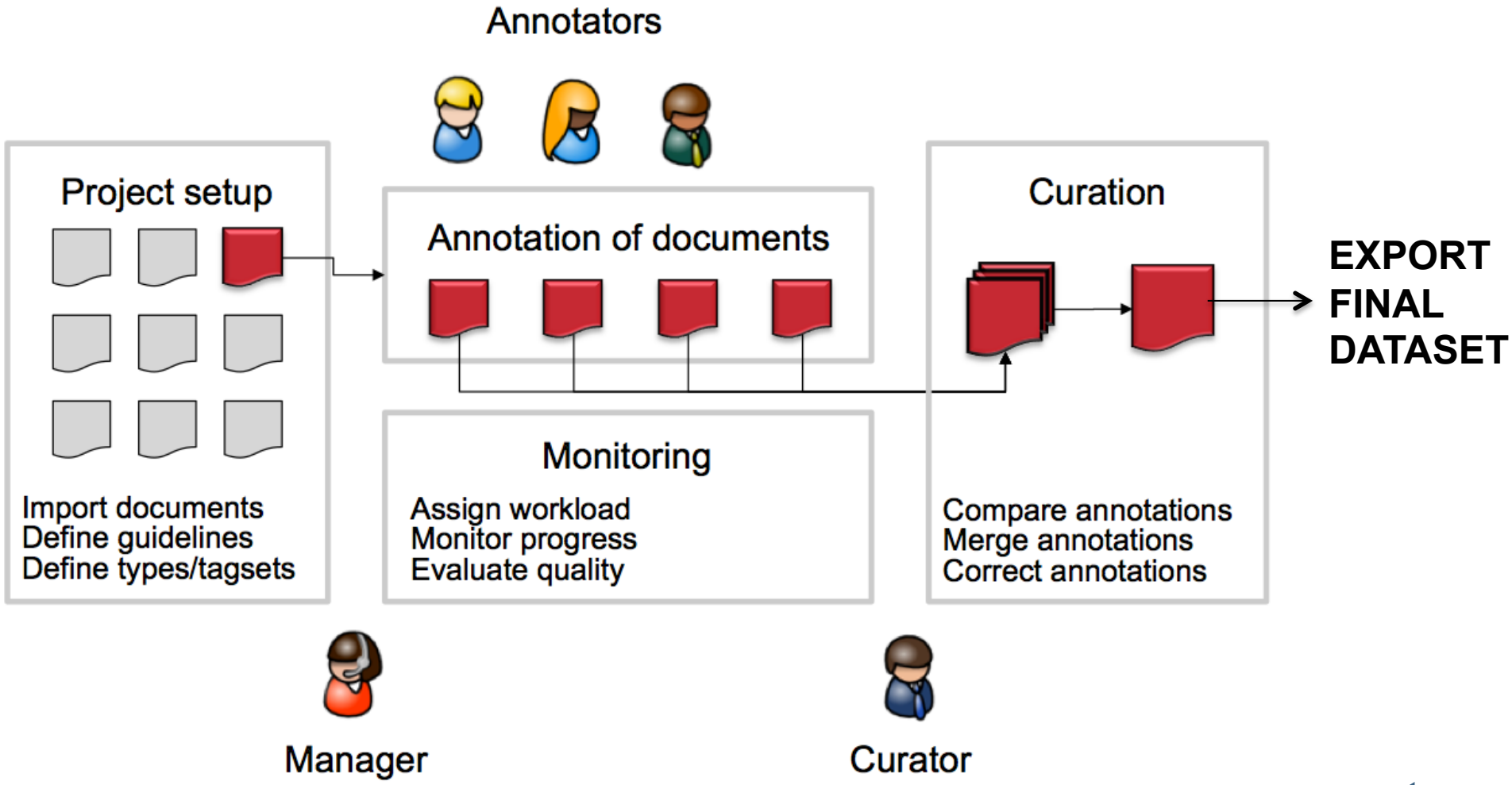

Projects

- Create new projects

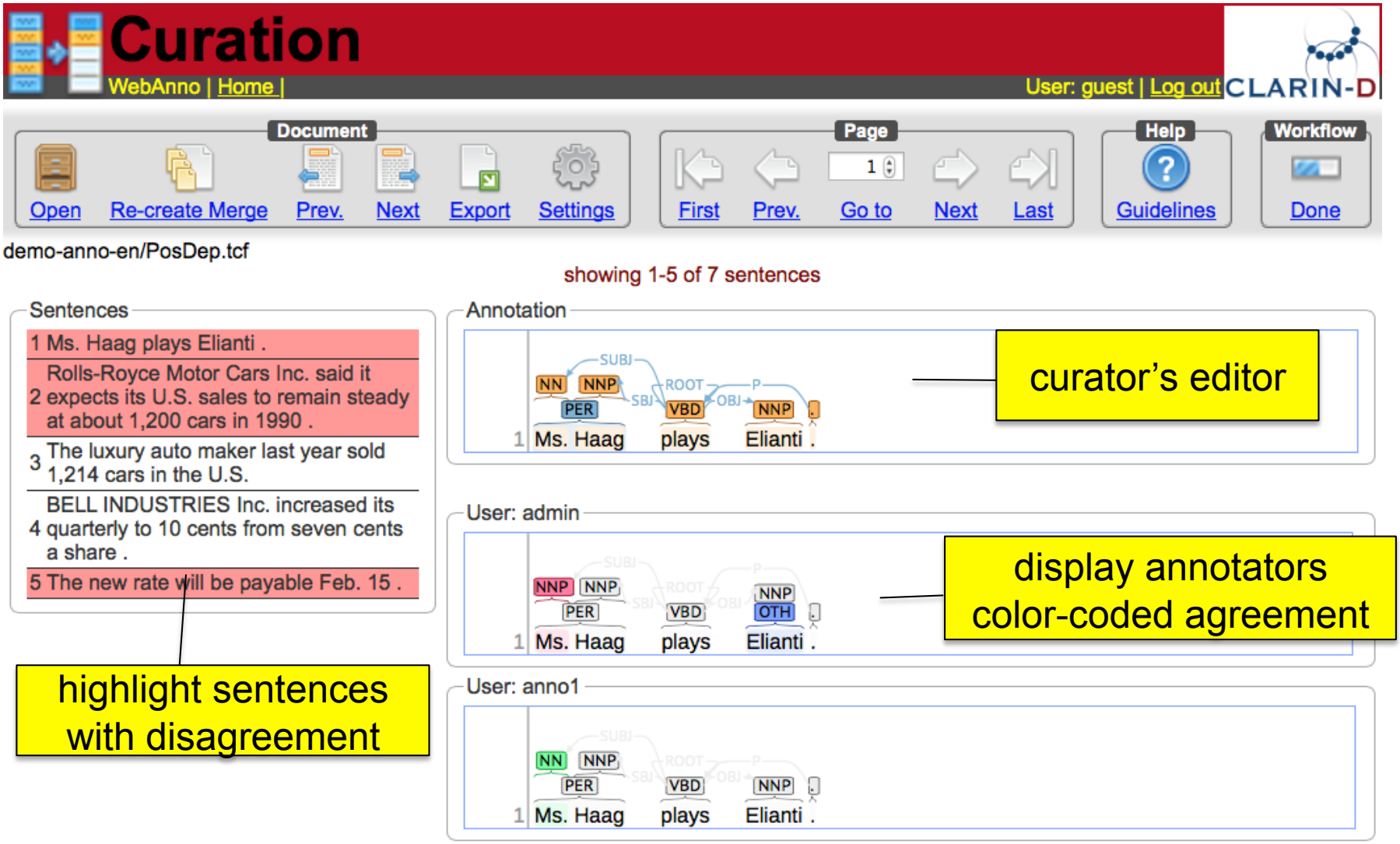

Manage users

- Create new user accounts

Workflow of a WebAnno project



Curation


Curation
WebAnno | Home |
User: guest | Log out
CLARIN-D
Document
Open
Re-create Merge
Prev.
Next
Export
Settings
Page
First
Prev.
1
Go to
Next
Last
Help
Guidelines
Workflow
Done
demo-anno-en/PosDep.tcf
showing 1-5 of 7 sentences
Sentences
1 Ms. Haag plays Elianti .
2 Rolls-Royce Motor Cars Inc. said it expects its U.S. sales to remain steady at about 1,200 cars in 1990 .
3 The luxury auto maker last year sold 1,214 cars in the U.S.
4 BELL INDUSTRIES Inc. increased its quarterly to 10 cents from seven cents a share .
5 The new rate will be payable Feb. 15 .
Annotation
User: admin
User: anno1
curator's editor
display annotators color-coded agreement
highlight sentences with disagreement

Built-in layers vs. custom layers

TECHNISCHE UNIVERSITÄT DARMSTADT

- **WebAnno offers various built-in annotation layers**
  - User can immediately start annotating
  - Only linguistic layers
  - Layer semantics are known

- **Custom layers allow WebAnno to be adapted to unforeseen tasks**
  - Adapt to non-linguistic annotation tasks
  - Adapt to unforeseen linguistic annotation tasks
  - Layer semantics are unknown

- **Import/export of annotated data**
  - Layers with known semantics convert from/to many formats (TCF, CoNLL, …)
  - Layers with unknown semantics convert from/to generic formats (XMI, …)

Layer types

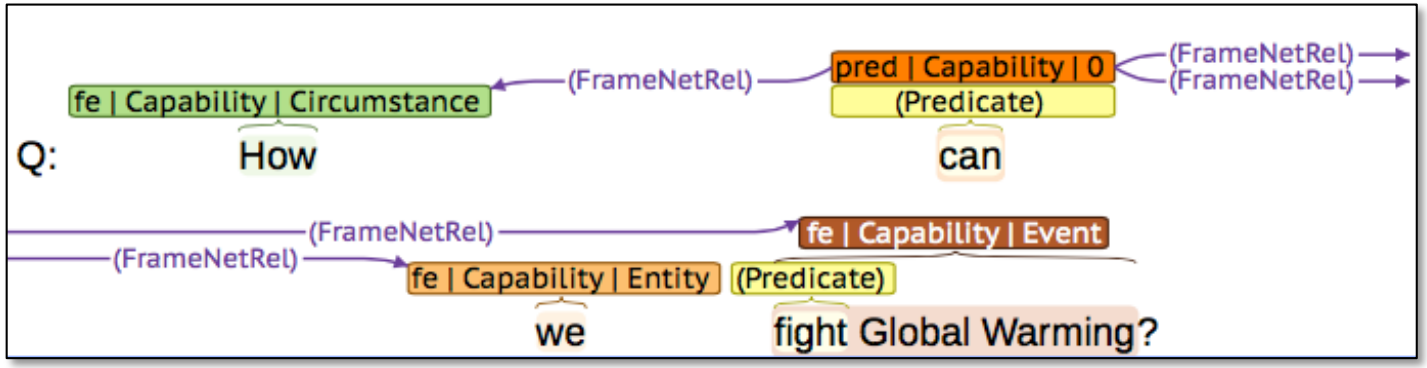
TECHNISCHE UNIVERSITÄT DARMSTADT

- **Existing built-in layers were generalized into three layer types**
- **Span layer** – POS, lemma, named entity, …
- **Relation layer** – Syntactic dependencies, …
  - Attaches to span annotations
  - Directed, reversible arcs
- **Chain layer** – Co-reference chains, …
  - Undirected arcs
- **Layers can be further customized using "behaviours"**
  - Character-based or token-based
  - Single/multiple token
  - Crossing of sentence boundaries
  - Stacking

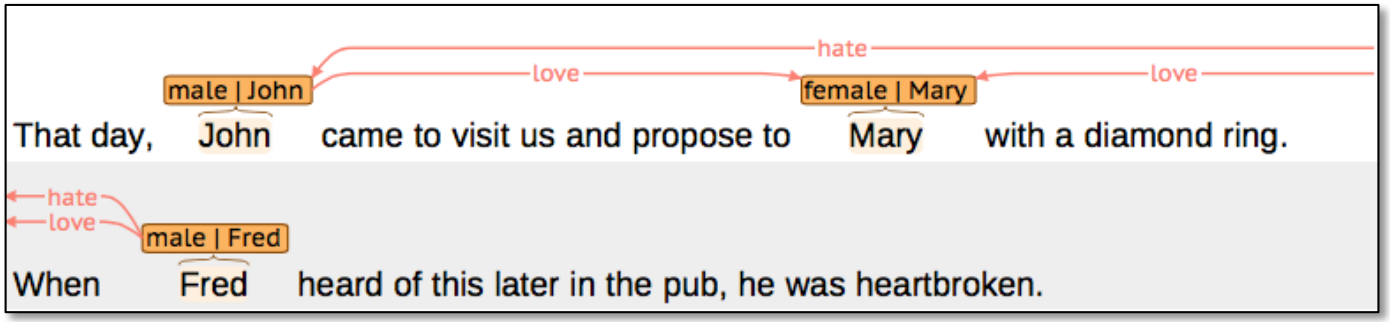
Custom layer examples

TECHNISCHE UNIVERSITÄT DARMSTADT

Semantic predicates and arguments (span/relation)



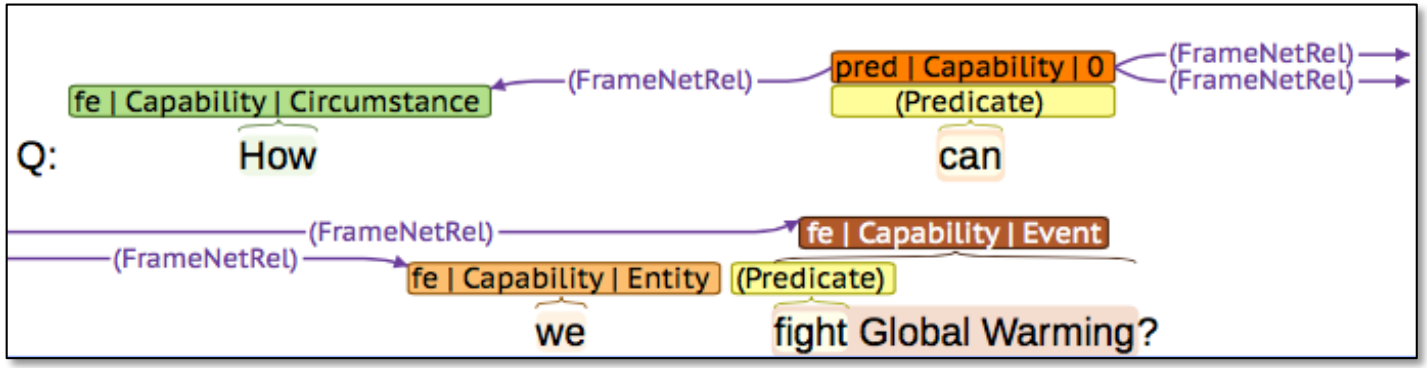
Person (span) / Relationship (relation)

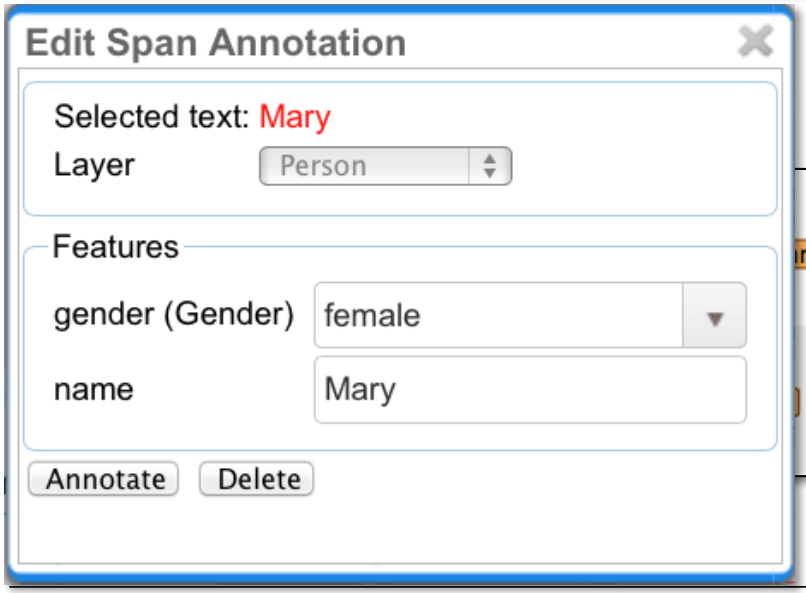


Custom layer examples

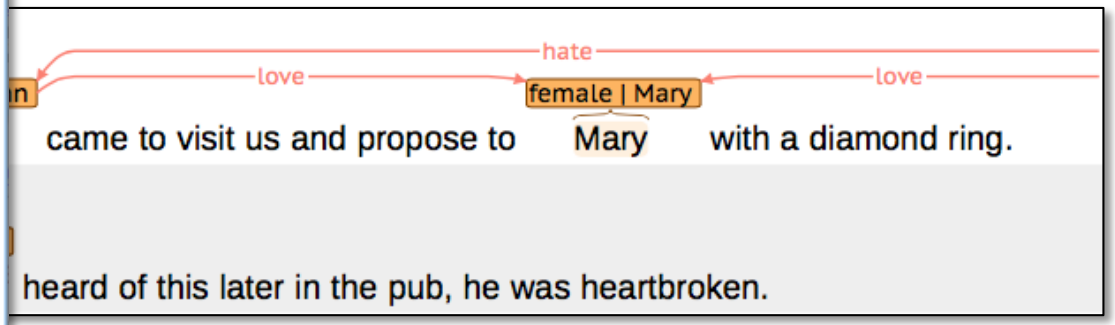
TECHNISCHE UNIVERSITÄT DARMSTADT

Semantic predicates and arguments (span/relation)

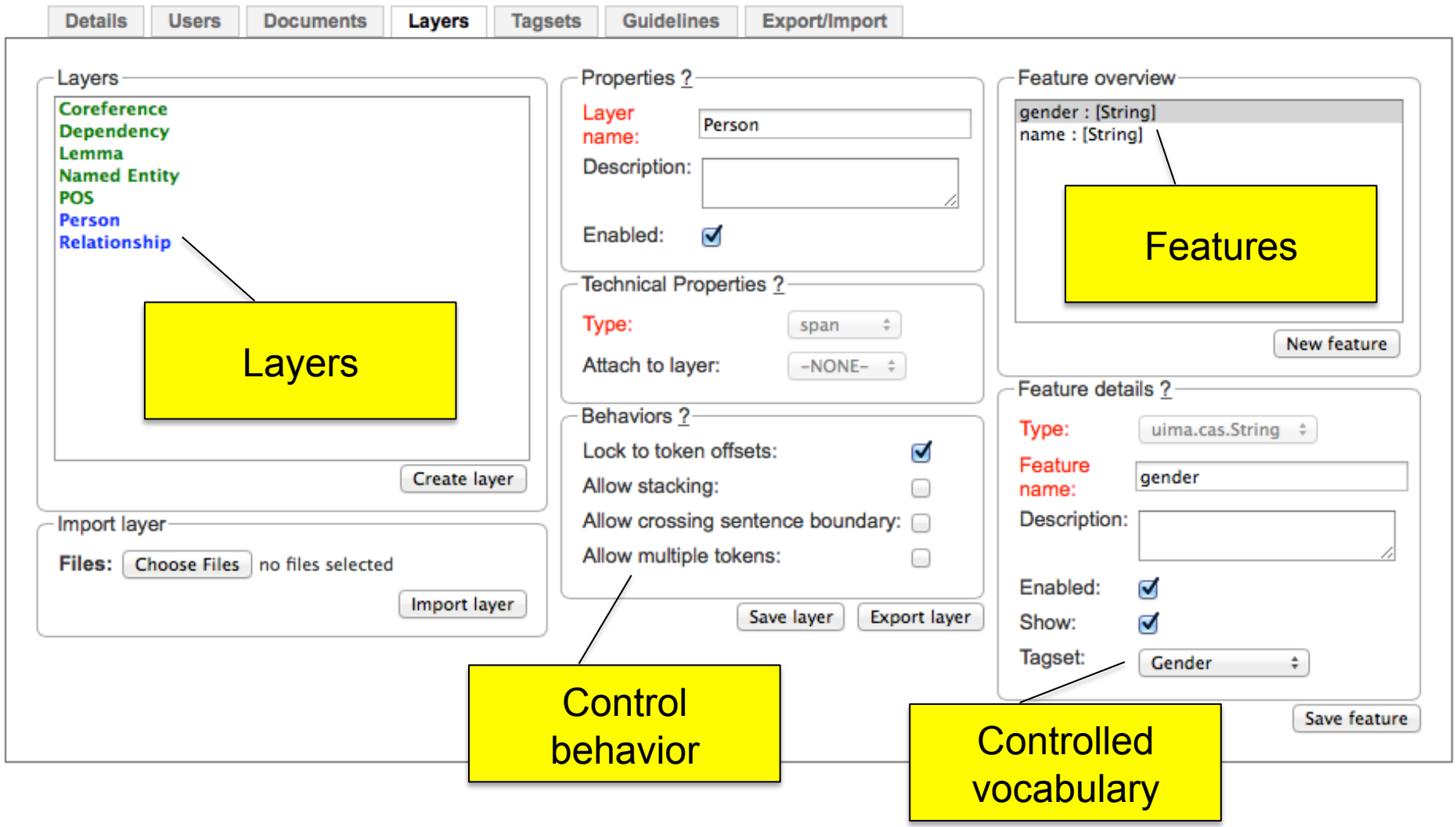




Person (span) / Relationship (relation)



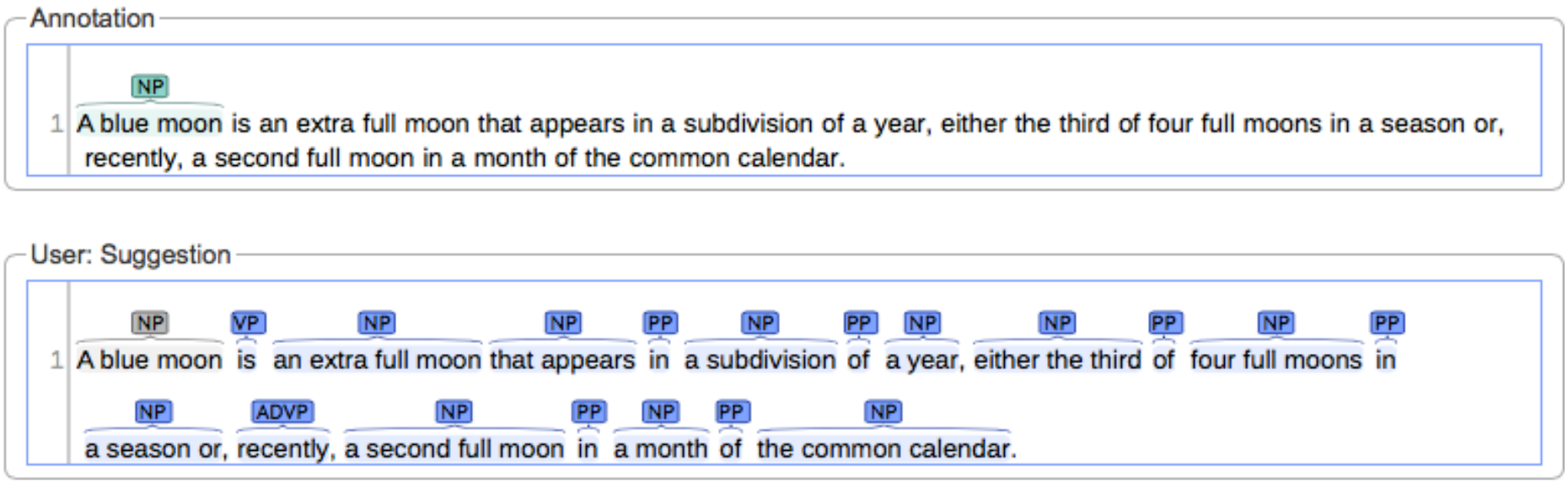
Custom layer configuration



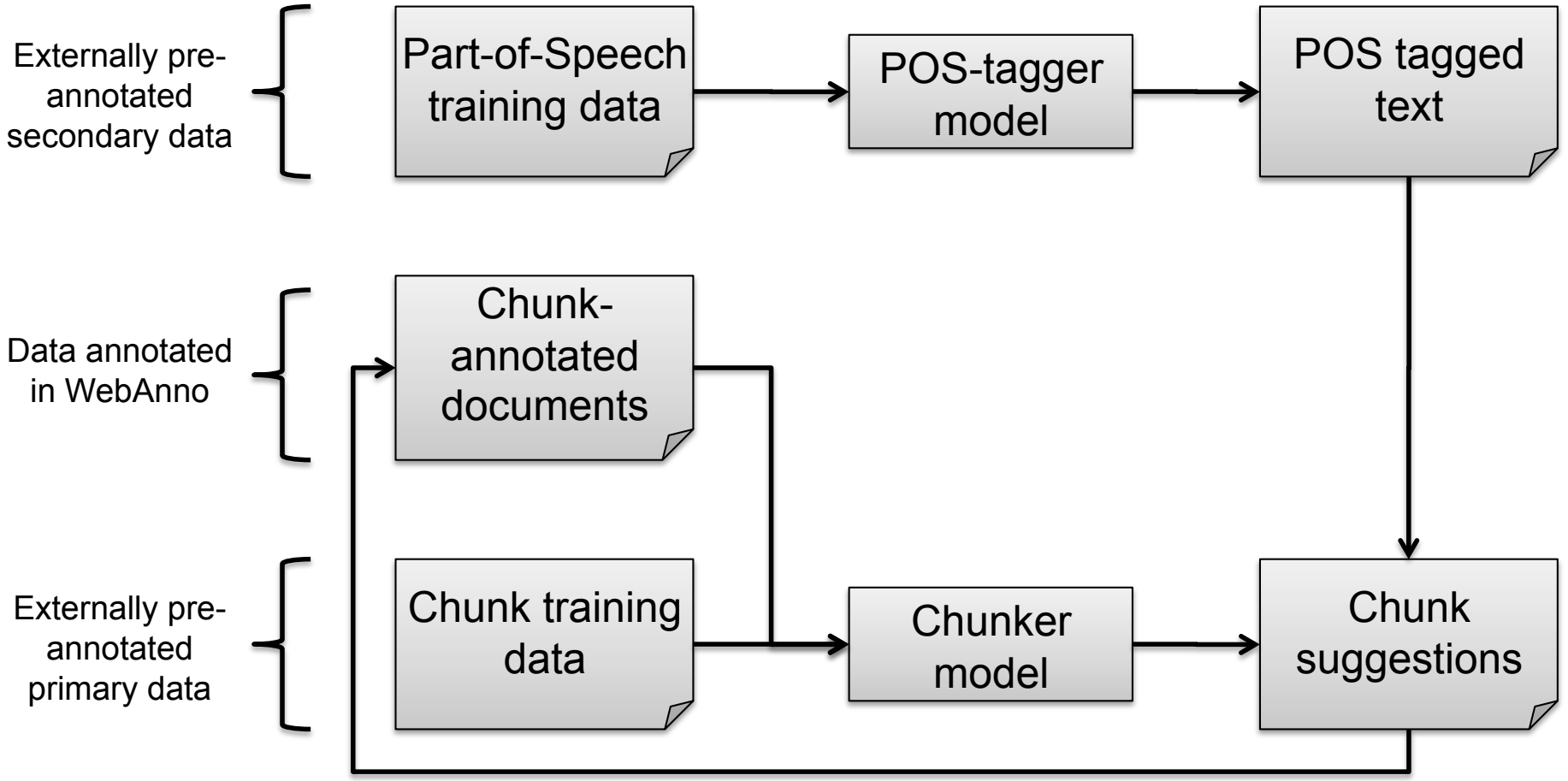
Integrated machine-learning

TECHNISCHE UNIVERSITÄT DARMSTADT

- Annotating data from scratch is more work than correcting
- WebAnno learns from pre-annotated data and makes suggestions
- Accept suggestions with a single click
- Correct suggestions to improve training data

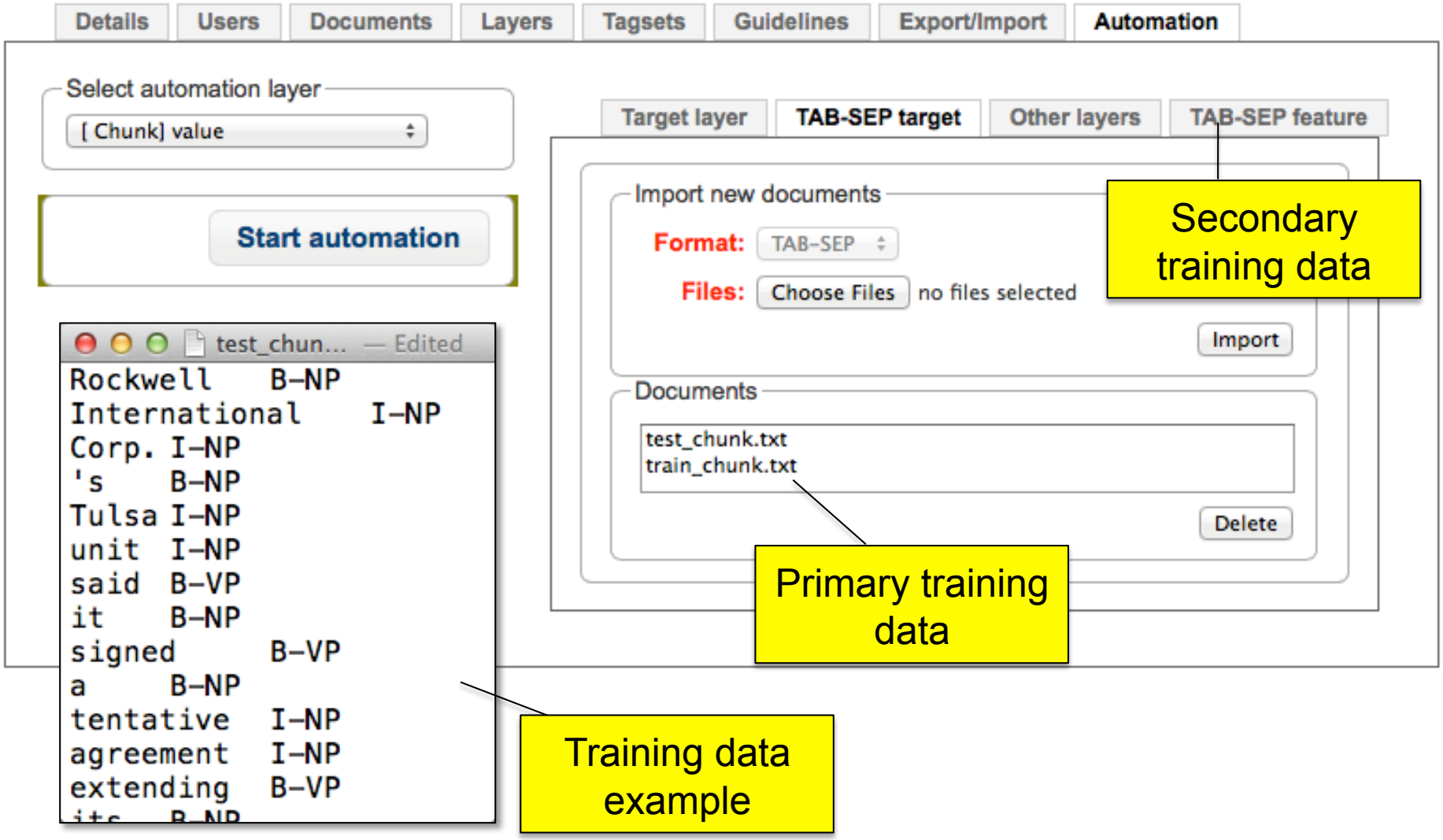


Example: Chunking



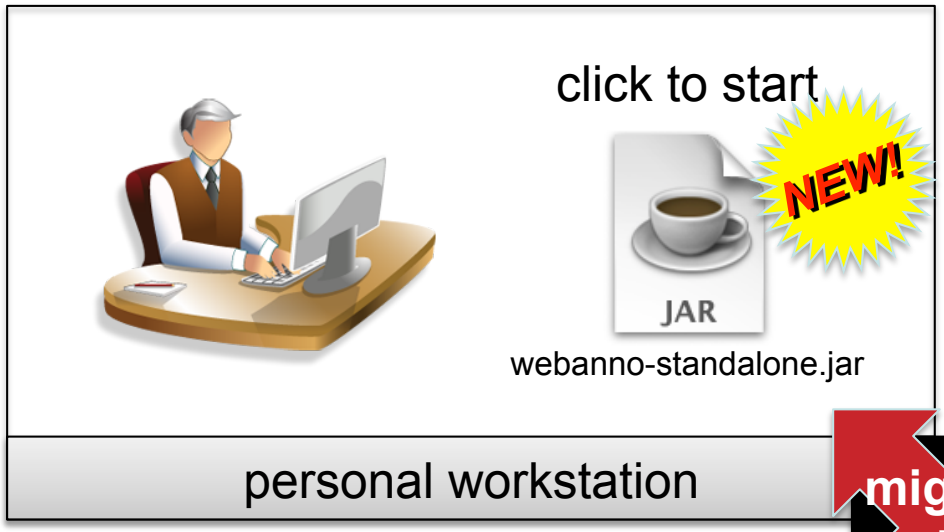
Automation configuration

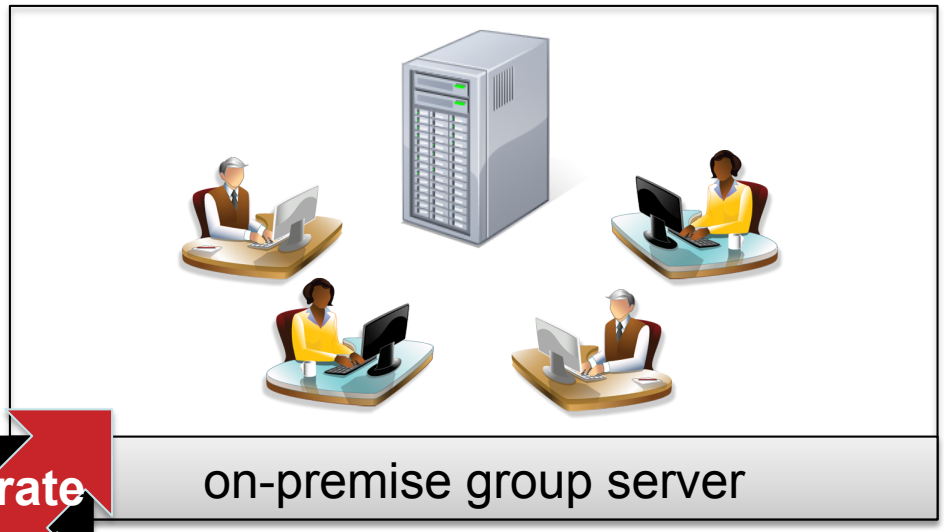
TECHNISCHE UNIVERSITÄT DARMSTADT



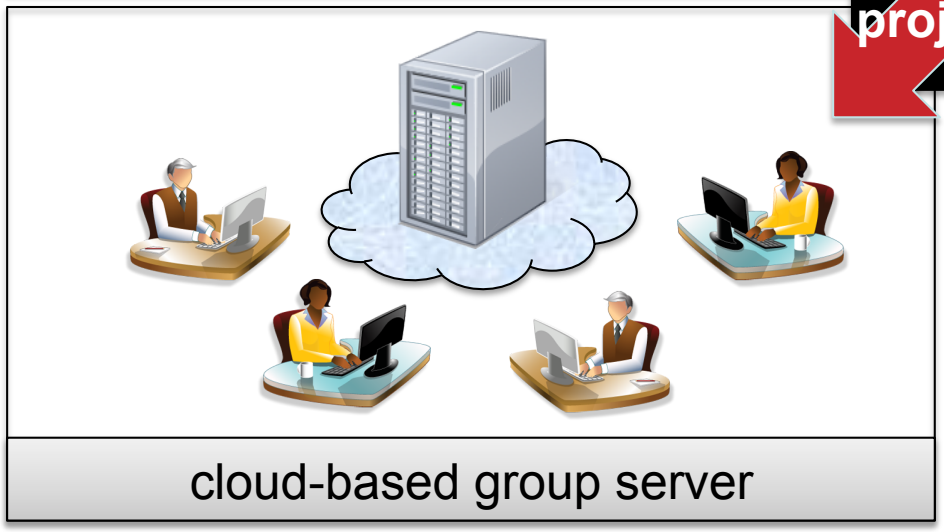
Deploy WebAnno as you need it

TECHNISCHE UNIVERSITÄT DARMSTADT





migrate projects

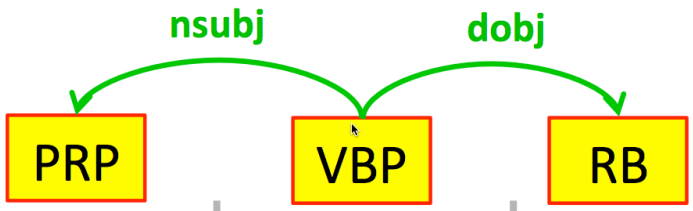


Where we want to go from here…

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Extend the scope of WebAnno**
  - Support for slot-based annotation layers (semantic annotations)
  - Tagset constraints
  - Support for more built-in linguistic layers
- **Improve continuously based on user feedback**
  - More efficient annotation interface
  - Support for additional corpus formats
  - … your feedback?
- **Deploy as a CLARIN infrastructure service**
  - CLARIN AAI support
  - Reduce administrative overhead for operators
  - Self-service for project managers

CLARIN-D



WebAnno

#WebAnno

TECHNISCHE UNIVERSITÄT DARMSTADT

Visit me in the demo session!

http://webanno.googlecode.com

WebAnno Community Survey

This survey is designed to help us understand how people use WebAnno or plan to use WebAnno. So if you are not already a user of WebAnno, we still appreciate your feedback. In this case, please answer the questions such as they reflect your intended use. If you find any of the questions not to be applicable, please simply leave it empty.

Thank your for participating in our survey!