

# The INCEpTION Platform: Machine-Assisted and Knowledge- Oriented Interactive Annotation



Jan-Christoph Klie, Michael Bugert, Beto Boudlosa, Richard Eckart de Castilho, Iryna Gurevych  
Ubiquitous Knowledge Processing Lab (UKP)  
Department of Computer Science, Technische Universität Darmstadt  
<https://www.ukp.tu-darmstadt.de>

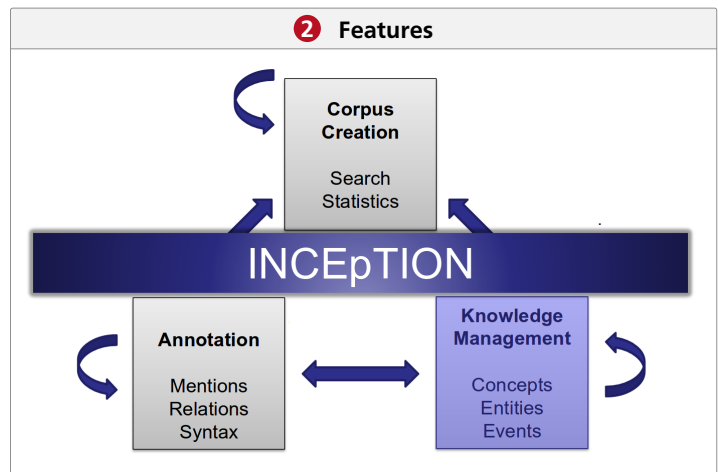


## 1 Motivation

- ▶ Many domains require annotated data (e.g. Machine Learning, Digital Humanities, Bioinformatics)
- ▶ Annotating texts is difficult and time consuming
- ▶ In the past, annotation tools were often rewritten from scratch for a specific task
- ▶ Knowledge management is often in a different tool
- ⇒ We develop INCEpTION, a generic and modular platform for interactive and knowledge-oriented annotation



## 2 Features



## 3 Annotation

- ▶ INCEpTION is a multi-user and -project annotation platform
- ▶ Fully supports span and relation annotations. Common annotation schemes are predefined (e.g. POS or NER), custom annotation schemes can easily be integrated
- ▶ Interactive annotation support via recommender is given to speed up the annotation process and improve quality
- ▶ Has active learning mode to improve the recommendations

PER Barack Obama Harvard Law School | ORG

In 1988 Obama enrolled in Harvard Law School, where he was the first black president of the Harvard Law Review.

## 4 Knowledge management

- Many annotation tasks require a knowledge base
- ▶ INCEpTION supports internal and external knowledge bases via SPARQL/RDF (e.g. WikiData or DBpedia)
- ▶ Built-in entity linking with smart recommendations
- ▶ Built-in fact linking for knowledge base population/completion
- ▶ Knowledge management is fully integrated into the platform

Barack Obama PER subject born in object Honolulu | LOC

Obama was born in 1961 in Honolulu, Hawaii, two years after the territory was admitted to the Union as the 50th state.

## 6 Extensibility

- INCEpTION is a modular annotation platform that supports a wide range of use cases. Standards are used where possible (e.g. UIMA, SPARQL, RDF)
- Simple customizability through extension points, e.g. for custom editors, additional recommenders, file formats, ...
- External recommender can be integrated in order to leverage existing, pre-trained models

## 5 Search

Text, annotations and knowledge base entries are fully indexed. They can be queried by the powerful MTAS Corpus Query Language.

<Named\_entity.value="LOC"/> Search

Wikipedia-Obama.txt

Senator from Illinois from

1961 in Honolulu

University in New York City in

organizer in Chicago

## 7 Conclusion

INCEpTION is –to the best of our knowledge– the first modular annotation platform which seamlessly incorporates recommendations, active learning, entity linking and knowledge management. Please ask us how we can accommodate your use case!

Text on GitHub



Code and data available at:  
<https://github.com/inception-project/inception>