

# Workshop INCEpTION

## Towards Interactive Semantic Annotation

March 12th (afternoon), March 13th (morning) 2018  
Lichtenberghaus, Dieburger Str. 241, 64287 Darmstadt



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



UBIQUITOUS  
KNOWLEDGE  
PROCESSING

---

## INCEpTION: Towards Interactive Semantic Annotation

---

End-to-end deep learning makes the need for semantically annotated text corpora larger than ever. We need smart, flexible, and reusable annotation tools to create annotated data at the lowest possible cost. Such tools need to target the semantic layer and support the identification and disambiguation of concepts and entities, their interactions on the textual and conceptual levels, and the circumstances of their interactions, allowing annotations for tasks such as cross-document event co-reference, semantic parsing, interactive knowledge-base population, and other emerging tasks.

The investigation of semantic phenomena in text is a very interactive and incremental task in which the researcher extracts knowledge and gains insight on specific phenomena from the corpus. To support such a process, functionalities that are currently scattered across different tools need to be integrated: corpus search, text annotation, and knowledge management. Such an integrated solution can provide a phenomenon-based access to corpora, e.g., in order to investigate specific topics, events or meanings.

The DFG project **INCEpTION** (Interactive distributed corpus exploration and annotation infrastructure for large corpora and knowledge-bases) at **UKP Lab** of the Technische Universität Darmstadt aims to:

- Enable phenomenon-based semantic annotation by incorporating corpus search, annotation, and knowledge management into a comprehensive annotation workbench.
- Optimize these functionalities by integrating state-of-the-art machine learning algorithms, e.g., for personalized clustering of search results, mention detection, or concept disambiguation.
- Offer a community-oriented, open, customizable, and extensible annotation platform that allows researchers to integrate their own corpora, machine learning algorithms and knowledge bases, and which is theory-agnostic and applicable to a wide range of text analysis tasks.

To align the development of the annotation platform with the community requirements and to make it effective in a wide range of semantic tasks and scenarios, the workshop will provide a forum for researchers who are in need of annotated data to exchange their expectations, requirements and experience. We will discuss current and future needs of interactive semantic annotation, addressing questions such as:

- What are the requirements resulting from current NLP research tasks, e.g., semantic parsing, interactive knowledge base construction, cross-document event co-reference, and where are the major bottlenecks of existing corpus annotation tools and workflows?
- How to allow researchers to automate processes in the platform, e.g., by integrating custom machine learning functionalities, such as active learning for cross-document annotation tasks, or pre-trained deep learning models?

- Which machine learning approaches are particularly well suited for human-in-the-loop scenarios (e.g., online learning, reinforcement learning, etc.), what are the current challenges in deploying them in practice, and how can these challenges be resolved?
- How to radically speed up the corpus annotation, e.g., by leveraging knowledge from similar tasks, user interaction data, other users, etc.?
- How to provide custom user interfaces, e.g., for annotating PDFs of scientific papers, highly-efficient crowdsourced annotation, multi-modal annotation, etc.?
- How to enable low-effort customizable data exchange between the researchers' usual tools and the annotation platform?

The workshop will provide a great opportunity to:

- Learn about the current status and plans of the INCEpTION project
- Present annotation use cases and requirements of your tasks
- Join the community of project users and supporters designing the annotation platform in a collaborative effort to enable novel research in large-scale semantic processing supported by interactive machine learning techniques.

The workshop will guide the evolution of the INCEpTION platform to meet the needs of the participants' current and future research.

---

## Programme:

---

### Monday, 12th March

12:00 - 12:30 Reception / Arrival

12:30 - 13:30 Lunch

#### \*\*\* Block 1 - Challenges \*\*\*

13:30 - 14:00 Welcome talk

14:00 - 15:30 Use case carousel

15:30 - 16:00 Coffee break

#### \*\*\* Block 2 - Methods \*\*\*

16:00 - 18:00 Machine-assisted  
annotation methods

### Tuesday, 13th March

#### \*\*\* Block 3 - INCEpTION \*\*\*

09:00 - 10:00 Project / tool presentation

10:00 - 10:30 Hands on demo

10:30 - 11:00 Coffee break

#### \*\*\* Block 4 - Wrap Up \*\*\*

11:00 - 12:00 Discussion rounds

12:00 - 13:00 Presentation and discussion  
of the results

13:00 - 14:00 Lunch

---

**Sign up until: February 28th, 2018**

For sign-up and questions, please contact

Beto Boullosa

boullosa@ukp.informatik.tu-darmstadt.de

Phone: +49 6151 1625299

---

For more detail on the programme and additional information, access  
the workshop homepage in the following link: <https://goo.gl/rCVruD>

