Overview of Potential Topics

UNI FREIBURG

T5. News Bias Detection

T6. How Fair are Researchers? Analyzing Biases of Paper Citations

T7. A Platform for Meta-Learning of Entity Linking

T5: News Bias Detection

- **News bias**: An article is defined as biased if the argumentation in the article is very partisan, i.e., if the article follows an agent (e.g., party or person) in a blind, prejudiced, or unreasonable manner.
- Document classification task (see SemEval 2019-Task 4):

  Determine whether a given news article is biased or not.

T5: News Bias Detection

- **Data set** given from SemEval 2019-Task 4:
  1 million English news articles labeled by the overall tendency of the publisher (biased/unbiased).

- What to do?
  → Develop & evaluate supervised machine learning approaches
  → Using **SVM** with hand-crafted features and/or using **deep learning**.
  - Examples of hand-crafted features: concerning writing style, news article topics, etc.
  - Examples of deep learning methods: recurrent neural network, convolutional neural network, convolutional recurrent neural networks (reusing existing implementations)

T5: News Bias Detection

Student‘s prerequisites:

- Interest and knowledge in data mining, binary classification, potentially also deep learning
- Knowledge in NLP/text mining a plus
- Knowledge in deep learning a plus

T5: News Bias Detection – Mandatory Task

- Build a news bias classifier based on simple features (sentiment lexicon or similar), using k-fold cross-validation.

T6. How Fair are Researchers? Analyzing Biases of Paper Citations

- Different kinds of **citing bias** possible in case of scientific publications.
- For instance,
  - **Content of written text**:
    - scientific results are only reported selectively.
    - publications are cited from specific disciplines and areas only (e.g., computer science, but not from psychology).
  - **Authors of citeable publications**:
    - unreasonable preference of self-citations and citations of papers authored by colleagues or friends.
  - **Source of citeable publications**:
    - preference of papers published at prestigious venues etc.

T6. How Fair are Researchers? Analyzing Biases of Paper Citations

- Task: **Derive lessons learned when citing bias occurs and in which ways**.
- Ultimate goal: reduce the citing bias using the lessons learned for citation recommendation.

→ Data mining & knowledge discovery, including natural language processing (information extraction on texts), correlation analysis, etc.

T6. How Fair are Researchers? Analyzing Biases of Paper Citations

UNI FREIBURG

- Using the Microsoft Academic Graph (MAG) as database.
- Information about authors, papers, conferences, citations, research fields, etc. available.
- Covering all scientific fields.

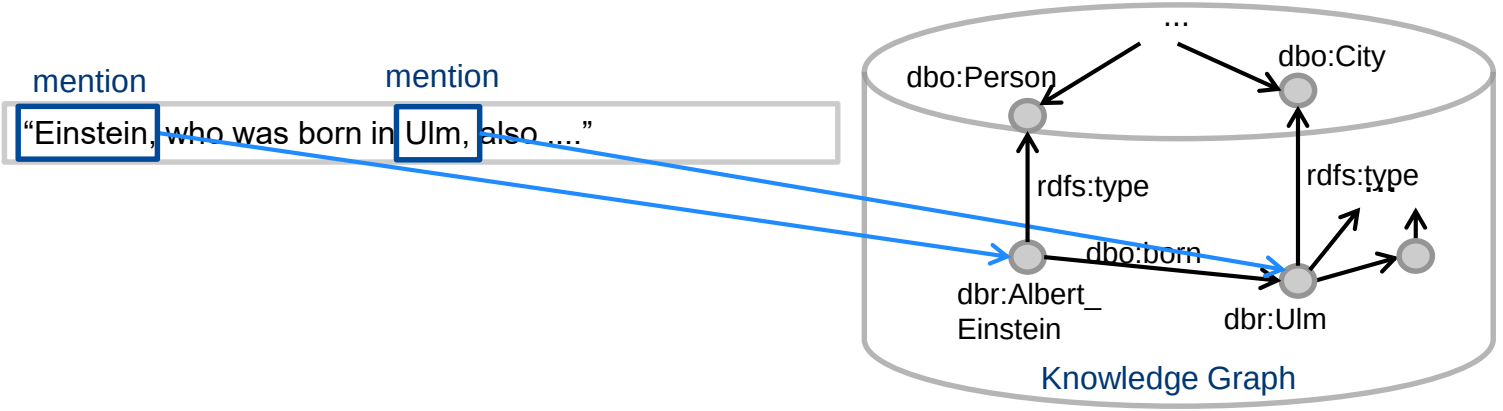
T6. How Fair are Researchers? Analyzing Biases of Paper Citations – Mandatory Task

UNI FREIBURG

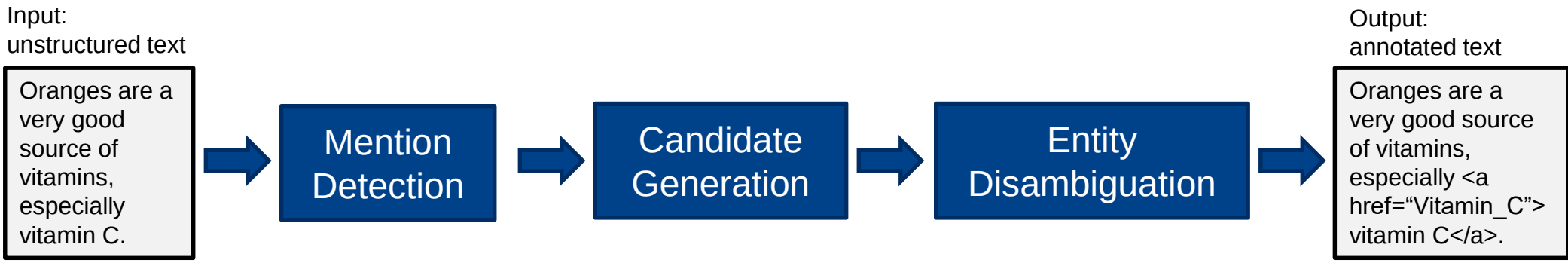
- Citing bias analysis w.r.t. self-citations (which authors/institutions/ venues/disciplines etc. have a strong tendency to cite themelves instead of others.)

T7. A Platform for Meta-Learning of Entity Linking – Entity Linking

- Entity linking:=linking phrases (“mentions”) in text to entities in a knowledge graph (KG; also called knowledge base, KB).
- KG as formal representation of knowledge.
  - KG contains entities, classes, relations, and attributes.
  - Here we consider only RDF KGs (Resource Description Framework Knowledge Graphs).
- Often focus on *named* entities (e.g., persons, organizations, locations, etc.).



T7. Steps of Entity Linking

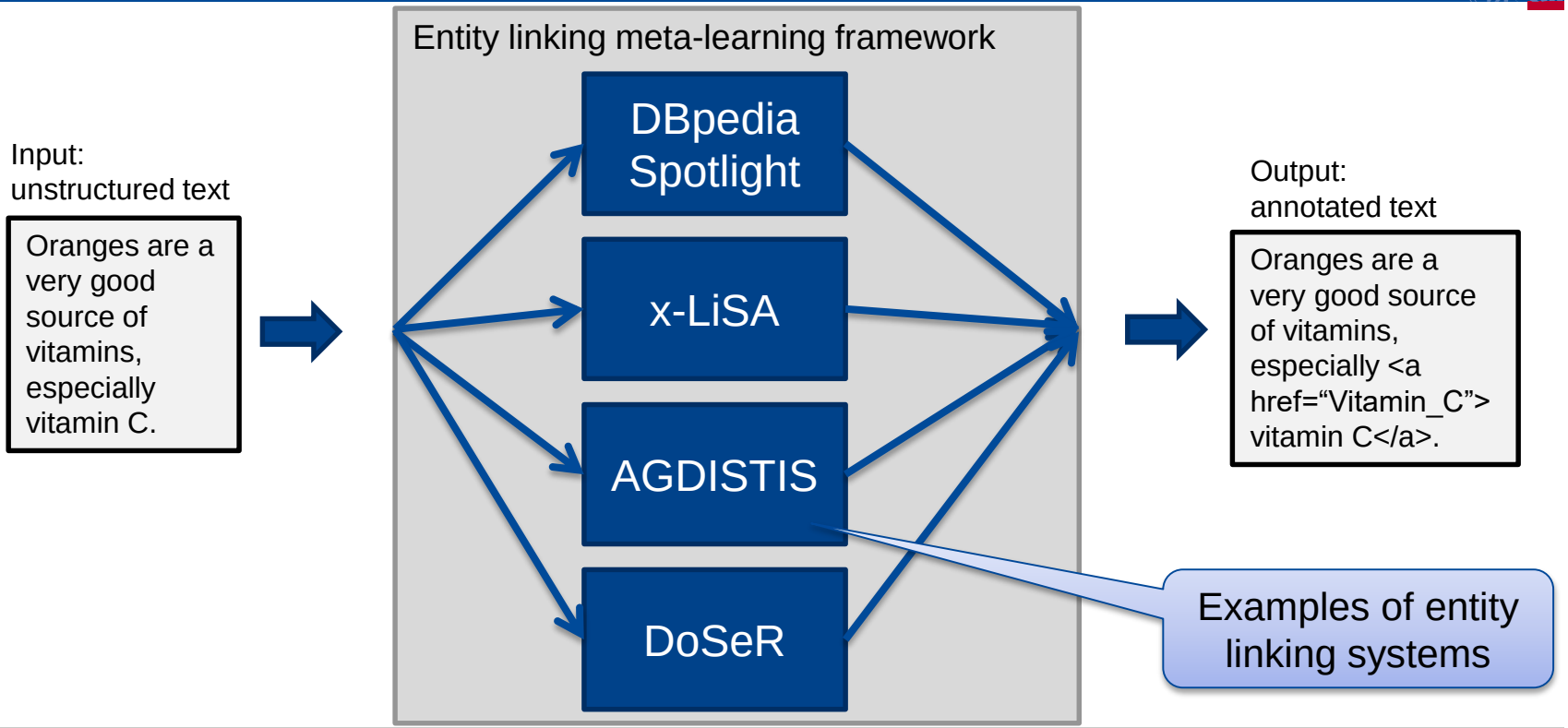


T7. A Platform for Meta-Learning of Entity Linking

UNI FREIBURG

- Overall goal: Learn when to use which entity linking system.
- Task: Build a platform which combines several entity linking systems:
  - Implement/redeploy several entity linking approaches.
  - Develop a platform which calls these entity linking approaches via APIs.
  - Apply a very basic meta-learning approach (majority voting, incl. cross-validation) for producing the “optimal” text annotations.
  - Implement a GUI for this system.
- Focus is not on developing meta-learning algorithms, but to develop an easy “plug & play” platform.

T7. A Platform for Meta-Learning of Entity Linking



T7. A Platform for Meta-Learning of Entity Linking – Mandatory Task

UNI FREIBURG

- Develop a very lightweight framework which can call 3 entity linking approaches via API calls.