Bachelorseminar

im Fachgebiet Deep Semantic Learning

(Seminar technische/praktische Informatik)

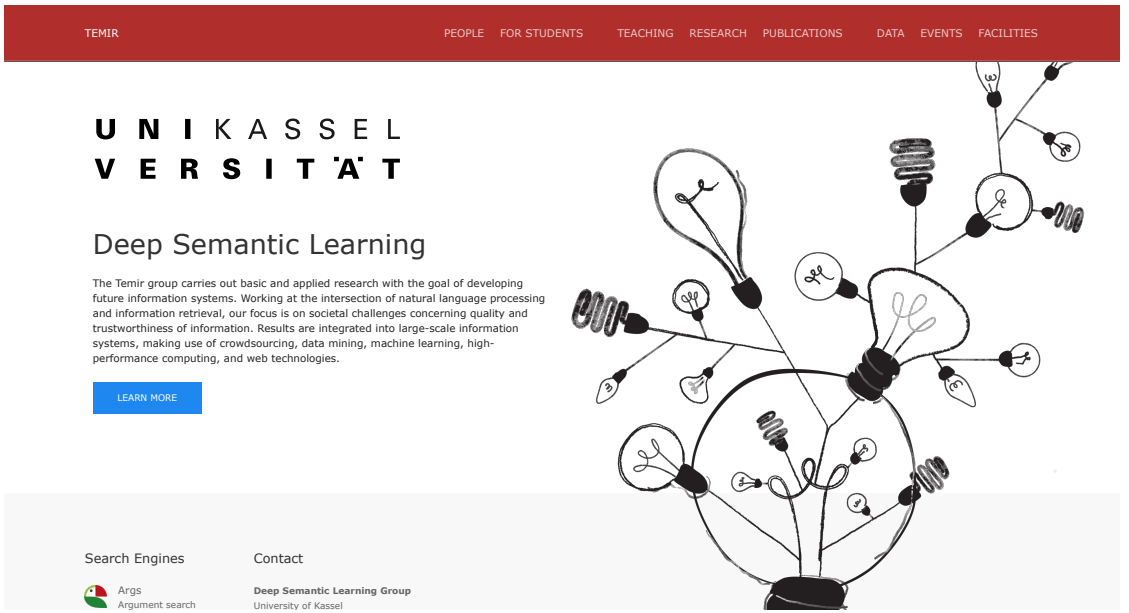
Summer term 2025

About us



Martin Potthast Tim Hagen Niklas Deckers

You can say “you” to us



[kassel.webis.de]

Your Task

- Conduct literature research for one of the following topics
- Write a term paper which summarizes the field
- Deliverables
  - Presentation
    (25 min presentation + 10 min Q&A)
  - Term paper
    (10 pages + unlimited space for references; LaTeX template provided)
- What makes a good seminar paper/presentation?
  - Identify and motivate a problem and describe its negative impacts
  - Provide illustrative examples to clarify the problem
  - Give a comprehensive overview on the existing literature to your peers
  - Clearly outline approaches to the problem and relate them to each other
    Don't just list approaches
  - Identify gaps in current research and potential future work

Context

- Topics are closely related to our research interests and cover a broad range
- Choose the topic that fits your own interests best
- Diving into a research field through this seminar can be a good starting point for future bachelor projects and bachelor theses with us
- Prerequisite: Foundational knowledge in machine learning
  e.g., as taught in our course “Machine Learning for Language Technologies” [temir.org]

Getting to Know Each Other

- What are you interested in?
- Have you had prior exposure to
  - machine learning,
  - natural language processing or
  - information retrieval?
- What is your scientific background?

Topic: Reasoning-Expectation in Prompts

Task

Collect all the prompts that have been used, for example, in papers at ECIR 202x or SIGIR 202x, and analyze how complex the reasoning is that the prompts expect the model to do. Also include research on the reasoning abilities of language models.

Supervisor: Martin Potthast (martin.potthast@uni-kassel.de)

Topic: Interpreting CLIP Space

Task

The CLIP space describes a relationship between images and text by embedding both modalities into a shared embedding space. However, the structure of this space is highly arbitrary and by far not optimal by many means. Approaches may optimize for aspects such as isotropy to obtain more robust embeddings. This and other approaches should be identified and compared. Applications in the field of generative models are of special interest for this seminar.

Supervisor: Niklas Deckers (niklas.deckers@uni-kassel.de)

Topic: Learning Distributions for Generative Models

Task

When using generative models to generate information such as images and text, a random distribution is used to sample from, especially to avoid generating identical information every time. With variational autoencoders (VAE) as the first generative models, distributions have been learned to resemble the provided datasets. This eventually resulted in the widespread application of image models like Stable Diffusion or other modalities. There are different strategies for learning a distribution from data that can be sampled from. These should be compared. An emphasis is on the use of different loss objectives. These approaches differ from learning a deterministic model.

Supervisor: Niklas Deckers (niklas.deckers@uni-kassel.de)

Topic: Topic Modelling on a Semantic Embedding Space

Task

Embedding spaces are designed with certain properties in mind, which are expressed in the form of losses during training. However, often a discrete modelling of the elements embedded into the space (and thus also the embedding space itself) is desired. Such discrete properties include class labels (classification), semantic clusters and their labels, and hierarchical structures. Bringing these properties back into the object space (text or images) is especially interesting to quickly assess the content, context, and extent of the structures on the embedding space.

Supervisor: Niklas Deckers (niklas.deckers@uni-kassel.de)

Topic: LLMs and Causality

Task

Can Large Language Models (LLMs) truly understand and reason about causality, or do they simply recite patterns from their training data? Explore this topic by examining methods that use LLMs to extract causal knowledge from text, methods that extract causal knowledge from LLMs, and methods used to evaluate their causal reasoning abilities. Also investigate the existence and quality of LLM-generated causal datasets, including their usefulness compared to human-curated datasets, and summarize key findings and ongoing debates in recent research on the capabilities and limitations of LLMs in causal tasks.

Supervisor: Tim Hagen (tim.hagen@uni-kassel.de)

Topic: Models for Causality Extraction

Task

Causality extraction is the task of identifying cause-effect pairs in natural language text. Even before large language models (LLMs), statistical, machine learning and deep learning models have been applied to this task. The question naturally arises whether LLMs outperform more traditional approaches. Research prior work on causality extraction and identify the key ideas of the proposed algorithms and architectures and relate them to each other.

Supervisor: Tim Hagen (tim.hagen@uni-kassel.de)

Topic: Causal Inference with Causal Graphs

Task

The results of causal extraction naturally form a causal graph where the nodes are abstract concepts or events and the edge relation is the causal relation. They can form the basis to causal inference and reasoning, for example, for causal question answering or root cause analysis. Research prior work on causal inference using causal graphs. Focus on pre-LLM approaches but also give an overview on different applications of LLMs to causal inference on causal graphs.

Supervisor: Tim Hagen (tim.hagen@uni-kassel.de)

What is Next?

- Give us your email address & join our Discord server (link via email)
- Until May 2 – Send us a ranked list of 3 topics that you would like to work on
- TBD – Tutorial session:
  - How to conduct literature research
  - How to manage a bibliography with LaTeX
  - How to do scientific writing
- TBD – Structure + Abstract check
  - Confirm with your supervisor that you are on the right track
- TBD – Halfway point
  - You should be halfway finished writing your term papers
  - Submit the current draft of your term paper anonymized to us
  - Everyone gets 3 other papers for review
- TBD – Tutorial session on scientific presenting
- TBD – Send us your slides for feedback (2 weeks prior to the presentation)
- TBD – Send us your final term papers and presentations
- TBD – Presentations