Multimodal Machine Learning Lab

Winter Semester 2025/2026

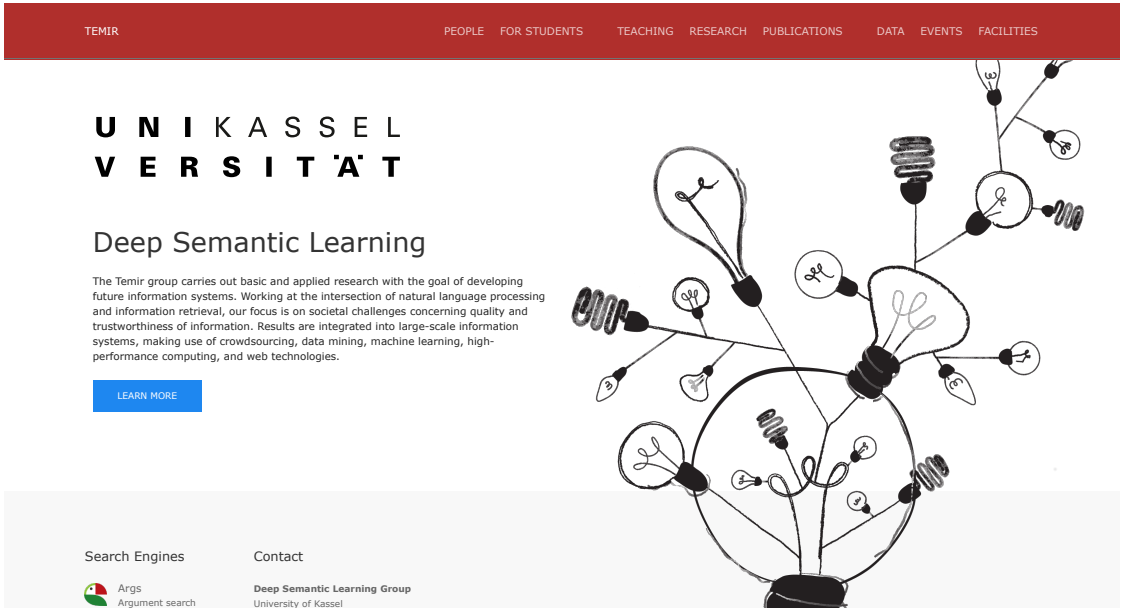
Niklas Deckers and Martin Potthast

Deep Semantic Learning
University of Kassel and hessian.AI

About us

Martin Potthast Niklas Deckers

You can say “you” to us



[kassel.webis.de]

Agenda

- Motivation
- Getting to Know Each Other
- Lab Organization and Research Objectives
- Technical Foundations
- Initial Task

Introduction

Guess the movie titles:

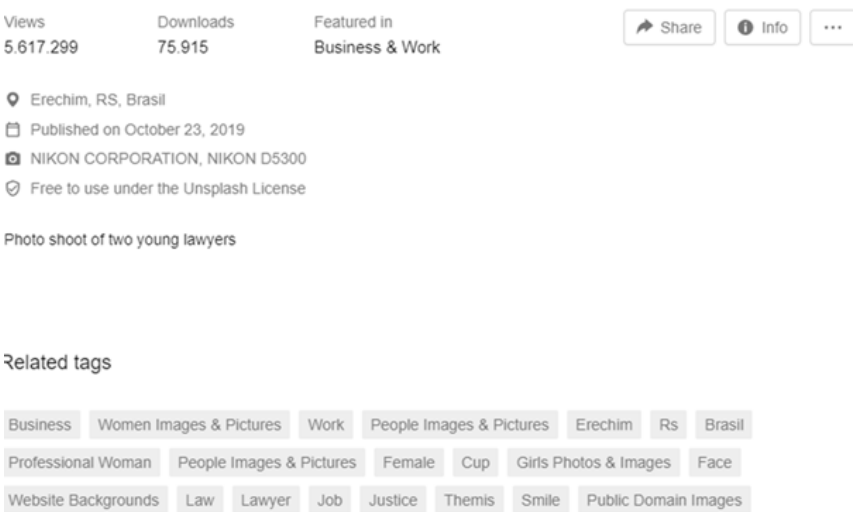
1. 
2. 
3. 

Introduction

- Stock images: Difference between literal description and underlying meaning

Introduction

- Stock images: Difference between literal description and underlying meaning
- Also applies to the assigned tags



Introduction

- Stock images: Difference between literal description and underlying meaning
- Also applies to the assigned tags
- Context is relevant

Eine Abkühlung ist bald vonnöten: In Deutschland steht die erste Hitzewelle des Jahres in den Startlöchern. © IMAGO / Shotshop

Introduction

How can the following concepts be represented using generative text-to-image models?

`friendship`

`diligence`

Introduction

How can the following concepts be represented using generative text-to-image models?

`friendship`

Introduction

How can the following concepts be represented using generative text-to-image models?

`diligence`

Is this really the best visual representation of these concepts?

Getting to Know Each Other

What experience do you have in the following subjects?

- Machine learning, deep learning

Getting to Know Each Other

What experience do you have in the following subjects?

- Machine learning, deep learning
- Language models like BERT

Getting to Know Each Other

What experience do you have in the following subjects?

- Machine learning, deep learning
- Language models like BERT
- Vision models, Unets, vision transformers

Getting to Know Each Other

What experience do you have in the following subjects?

- Machine learning, deep learning
- Language models like BERT
- Vision models, Unets, vision transformers
- CLIP

Getting to Know Each Other

What experience do you have in the following subjects?

- Machine learning, deep learning
- Language models like BERT
- Vision models, Unets, vision transformers
- CLIP
- Programming in Python

Getting to Know Each Other

What experience do you have in the following subjects?

- Machine learning, deep learning
- Language models like BERT
- Vision models, Unets, vision transformers
- CLIP
- Programming in Python
- PyTorch

Getting to Know Each Other

What experience do you have in the following subjects?

- Machine learning, deep learning
- Language models like BERT
- Vision models, Unets, vision transformers
- CLIP
- Programming in Python
- PyTorch
- Git, SSH, Slurm

Getting to Know Each Other

What experience do you have in the following subjects?

- Machine learning, deep learning
- Language models like BERT
- Vision models, Unets, vision transformers
- CLIP
- Programming in Python
- PyTorch
- Git, SSH, Slurm
- Text-to-image models like Stable Diffusion

Getting to Know Each Other

What experience do you have in the following subjects?

- Machine learning, deep learning
- Language models like BERT
- Vision models, Unets, vision transformers
- CLIP
- Programming in Python
- PyTorch
- Git, SSH, Slurm
- Text-to-image models like Stable Diffusion
- Prompt engineering

Lab Organization

- Weekly consultations (additional fixed time slot? where?)
- At the end of the semester: Written report (in groups) and presentation

Learning Objectives

- Work in a structured and self-supervised manner
- Work on a project of a larger scope
- Deal with open-ended tasks
- Groupwork and communication skills
- Apply and extend current research and tools in the field of multimodal models
- Develop and carry out experiments
- Scientific writing
- Apply machine learning to a real life problem

We would like to do real research with you!

Research Objectives for This Lab

- What can we learn about human–human and human–machine communication by using multimodal concepts to isolate the underlying meanings?
- What are the boundaries of this communication?
- Objective: Developing an application (web UI or game) that sets specific requirements to a communication scenario, and conducting experiments using this application

Outline of This Course

- Technical foundations
- Cluster onboarding
- Initial task: Deriving meaning from context
- Semiotics: Signs and their usage
- Communication scenarios: Real-life applications and artificially restricted scenarios
- Games as experiments
- User studies incorporating multimodality

Recap: Previous Course

- Motivation [temir.org]
- Demo [webis.de]

Foundations

- Embeddings [mlvu.github.io]
- Word2Vec [temir.org]
- Transformers [temir.org]
- CLIP and generative models [webis.de]

Initial Task

- Is there a natural order to emojis?
- How can the semantic meaning be inferred from emojis?
- Write a software that derives such a natural order
- Five-minute presentations (each student individually) on 05.11.2025: Approach, results, visualizations
- Discussion: What ingredients are needed? What problems may arise?