Deep Learning & Digital Humanities
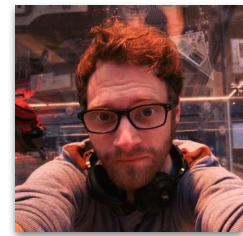
Kickoff

Dr. Steffen Eger*
Thomas Haider, M.A.**

* Natural Language Learning Group
** Max Planck Institute for Empirical Aesthetics
** IMS, University of Stuttgart

Agenda

- Our backgrounds
- Seminar goals

- Organization
  - Presentation
  - Jury
  - Term paper
  - Your grade
- Seminar topics & schedule
- Important links

Our backgrounds

Steffen Eger

- PhD in economics
- PostDocs in NLP 2014-2016, 2016-2018
- Independent Research Group Leader ("Nachwuchsgruppenleiter") since 2019
- Interested in Deep Learning for NLP
  - Humanities applications like poetry generation + semantic change
  - Also cross-linguality & representation learning

Our backgrounds

Steffen Eger

- Teaching

  Deep Learning for Natural Language Processing (each SS)

  https://www.informatik.tu-darmstadt.de/ukp/teaching_ukp/courses_3/previous_terms/sose_2019/deep_learning_for_natural_language_processing_2/inhalt_mit_marginalienspalte_75.en.jsp

- Links

  - IRG https://www.informatik.tu-darmstadt.de/aiphes/aiphes/irg_position/index.de.jsp

    (theses available)

  - Website

    https://www.informatik.tu-darmstadt.de/aiphes/aiphes/people_7/mitarbeiter_4_detailseite_72000.en.jsp

Seminar goals

The focus of the seminar is on **humanities applications** like *poetry generation and analysis, metaphor and emotion identification,* etc., and how to **solve these problems with Deep Learning**. *Students will read papers and present them during the seminar.*

Seminar goals

- The seminar is not an introduction to Deep Learning
- We will also not cover any programming for Deep Learning
- If you lack skills in Deep Learning, you need to be able to acquire them yourselves, mostly
  - at least the basics

- Caution: DL requires understanding of math (at least a little) and programming (at least a little)

Organization

The seminar takes places

- Thursdays from 18:15 to 19:45

- About twice per month

- In S1|08 18

- Find the exact schedule here:

  - https://github.com/SteffenEger/dldh/blob/master/index.md

Organization

To pass the course, you need to
- Present
  - around 15min
  - typically 1-2 papers

Organization

To pass the course, you need to

Session	
TOPIC 1 Talk 1 (15 mins) Talk 2 (15 mins) Discussion	**TOPIC 2** Talk 1 (15 mins) Talk 2 (15 mins) Discussion

1. Give a talk (15 mins)
   a. Max 2 people per talk
   b. Max 2 papers per talk
   c. Papers will be given only after consultation with us
2. Ask questions as a jury member
   a. Questions for each talk: understanding, like, dislike
   b. 3 -- 5 jury members
3. Write a term paper
   a. 4 -- 8 pages

Organization

To pass the course, you need to

- Give a talk (15 mins)
  - Inform your peers/us on time if you want to drop the seminar / talk
  - Prepare back-up slides

Organization

To pass the course, you need to

- Write a term paper
  - 4-8 pages
  - Either "**close reading**":
    - Discuss 1-2 papers in depth
  - Or "**distant reading**":
    - Give an overview of a topic, discuss 3-6 papers
  - **Discuss your topic with us before starting to write**
  - You may describe small-scale experiments
    - 2-3 pages
    - **Discuss with us**

Organization

Plagiarism

- Will lead to failing the course
- Plus you will receive a "note" at TU Darmstadt

Source: https://www.123rf.com/photo_36611606_bad-decision-red-rubber-stamp-over-a-white-background-.html

Organization

All of the following are considered plagiarism

- turning in someone else's work as your own

- copying words or ideas from someone else without giving credit

- failing to put a quotation in quotation marks

- giving incorrect information about the source of a quotation

- changing words but copying the sentence structure of a source without giving credit

- copying so many words or ideas from a source that it makes up the majority of your work, whether you give credit or not (see our section on "fair use" rules)

Source: https://www.plagiarism.org/article/what-is-plagiarism

Organization

Your grade
- 70% * term_paper + 30% * (presentation+jury+attendance)

- **Term paper** should be *clear, follow conventions of scientific articles, well organized, grammatically correct*, etc.
- **Presentation** should be *clear, interesting, entertaining, with pictures*, etc. Should be able to *answer* questions
- As a **jury** member, you should ask *meaningful questions*
- You should **attend** *regularly*

How a talk should be structured

1. Motivation (Research Question)
2. Method / Approach (Deep Learning?)
3. Resources / Data (Corpora?)
4. Experiments / Results
   a. How useful are the results for the humanities?
   b. How useful is the method for this problem?
   c. If the method is not DL: How could it be modelled differently?
5. Conclusion / Summary

Seminar schedule (tentative)

Session	Date	Topics	Presenters
1	24.10.	DH	Thomas Haider
2	31.10.	DL	Steffen Eger
3	14.11.	Corpora & Annotation	
4	28.11.	Spelling Normalization & OCR	
5	12.12.	Metaphors & Emotions	
6	19.12.	Semantic Change & Drift	
7	16.01.	Stylometry & Variation	
8	**23.01.**	Fiction & Narration	
9	30.01.	Poetry & Arts	

Seminar topics

Data & Agreements

- How to obtain labeled data (for humanity applications)?
  - Via experts or crowds
- How to ensure/measure the annotations are consistent?

DL

DH
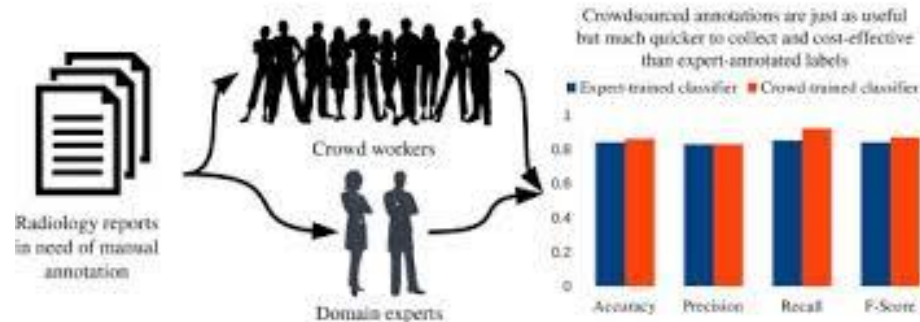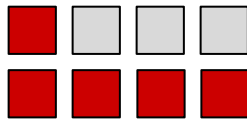
Semantic change & Semantic Drift?

- How to measure diachronic meaning change?
- How to quantify spatial meaning variation?

DL

DH

Seminar topics
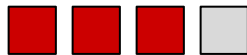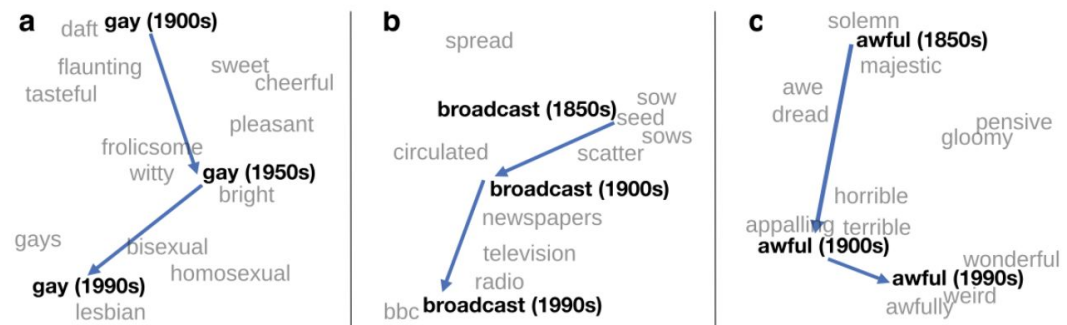
Metaphors & Emotions

- How to identify metaphors in texts?
- How to identify emotions?

DL

DH
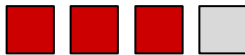
Seminar topics

Poetry & Fiction

- How to automatically generate poetry?

- How to identify relations among characters in novels?

DL

DH

Seminar topics
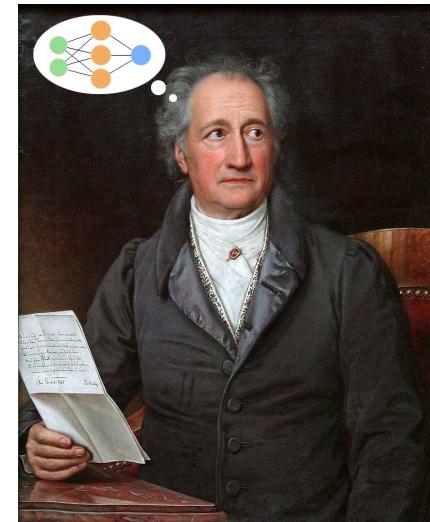
Spelling Normalization & OCR post-correction

- How to unify historical spelling variants? (*daß* vs.*dass*)
- How to post-correct OCR scans?
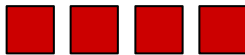
DL

DH

Seminar topics

Stylometry & Author identification

- How to identify the style of an author or transform styles?
- How to use style to attribute authorship?

DL
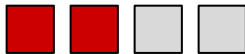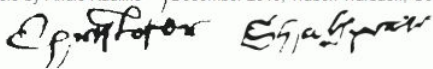
DH

Two Identities : One Man ?
Was "William Shakespeare" actually an alias name for Christopher Marlowe ?

Analysis by Andis Kaulins   December 2010, Traben-Trarbach, Germany

Portrait and Signature of Christopher (Marlowe) left, first name only
Portraits and Signature of Shakespeare (William) right, last name only

Note especially the similarity of the initial C and S and the initial H

Useful links

Moodle

- https://moodle.informatik.tu-darmstadt.de/course/view.php?id=789

Seminar schedule

- https://github.com/SteffenEger/dldh/blob/master/index.md

Email addresses

- eger@aiphes.tu-darmstadt.de

- thomas.haider@ae.mpg.de

- Always contact both of us

Deep Learning Courses

- DL4NLP2019 TUDA is on moodle

- Stanford DL4NLP: http://web.stanford.edu/class/cs224n/

Useful links

Template for term paper

- ● Will be provided

Template for presentation

- ● Will be provided

Your todos

- Decide on a topic
- Register for the topic (link will be available tomorrow)
  - Until next Thursday
- Register for a jury membership
- If there are more participants than talks:
  - You will be in the queue
  - We will try to organize extra sessions. Possible topics:
    - Deep Learning Background
    - Fiction & Narration
    - Suggestions?

THÄNK$!

References

- Zhao, Wei ; Peyrard, Maxime ; Liu, Fei ; Gao, Yang ; Meyer, Christian M. ; Eger, Steffen (2019):MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance., EMNLP 2019
- Gao, Yang ; Eger, Steffen ; Kuznetsov, Ilia ; Gurevych, Iryna ; Miyao, Yusuke (2019): Does My Rebuttal Matter? Insights from a Major NLP Conference. NAACL 2019
- Eger, Steffen ; Şahin, Gözde Gül ; Rücklé, Andreas ; Lee, Ji-Ung ; Schulz, Claudia ; Mesgar, Mohsen ; Swarnkar, Krishnkant ; Simpson, Edwin ; Gurevych, Iryna (2019): Text Processing Like Humans Do: Visually Attacking and Shielding NLP Systems. NAACL 2019
- Eger, Steffen ; Youssef, Paul ; Gurevych, Iryna (2018): Is it Time to Swish? Comparing Deep Learning Activation Functions Across NLP tasks. EMNLP 2018
- Eger, Steffen ; Daxenberger, Johannes ; Stab, Christian ; Gurevych, Iryna (2018): Cross-lingual Argumentation Mining: Machine Translation (and a bit of Projection) is All You Need! COLING18
- Schulz, Claudia ; Eger, Steffen ; Daxenberger, Johannes ; Kahse, Tobias ; Gurevych, Iryna (2018): Multi-Task Learning for Argumentation Mining in Low-Resource Settings. NAACL 2018
- Daxenberger, Johannes ; Eger, Steffen ; Habernal, Ivan ; Stab, Christian ; Gurevych, Iryna (2017): What is the Essence of a Claim? Cross-Domain Claim Identification. EMNLP2018
- Eger, Steffen ; Daxenberger, Johannes ; Gurevych, Iryna (2017): Neural End-to-End Learning for Computational Argumentation Mining. ACL 2017