Computational Linguistics

Dietrich Klakow & Stefan Thater

FR 4.7 Allgemeine Linguistik (Computerlinguistik)
Universität des Saarlandes
Summer 2013

This Lecture

- **Lectures:**
  - Tuesday, 14:15 – 15:45
- **Exercise sessions:**
  - Thursday, 14:15 – 15:15
- **Lecturers:**
  - Stefan Thater (Week 1–7)
  - Dietrich Klakow (Week 8–12)

Course website

- Lecture slides and exercise sheets can be obtained from the course website:
  - **http://www.coli.uni-saarland/courses/CL/2013**
- Please subscribe to the follow mailing list:
  - **http://ml.coli.uni-saarland.de/cgi-bin/mailman/listinfo/compling**

Exercise sheets

- Assignments will be given after every lecture
  - You will (usually) have 1 week to submit your solutions
  - Exercise sheets will usually consists of theoreticall questions as well as programming tasks
  - You have to get at least 50% of the points to be admitted to the final exam
- Programming tasks:
  - we recommend (and support) Python
  - consult with us if you prefer another language

Final Exam

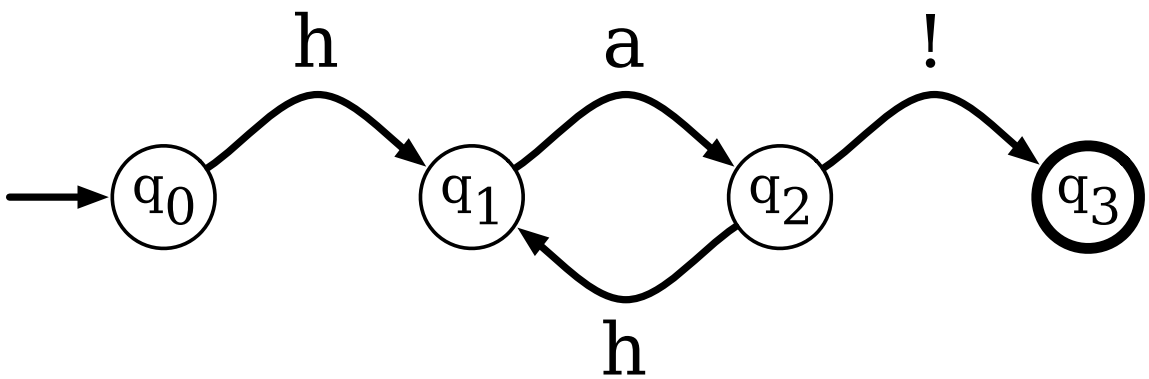
- The final exam will take place
  - Tuesday, July 23rd or July 16th
- Registration deadline: **July 8th**
- Note: Dates to be confirmed!

Textbooks

- D. Jurafsky and J. H. Martin: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall, 2009.
- C. Manning and H. Schütze: Foundations of Statistical Natural Language Processing, MIT Press, 1999.

Lecture 1 (Thater)
Finite State Automata

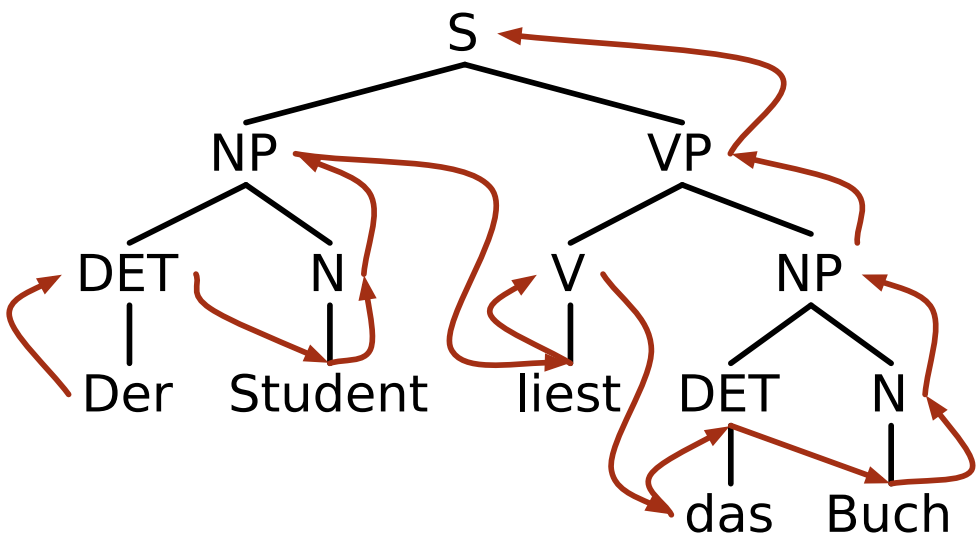
- Deterministic & non-deterministic finite state automata
- Algorithms for recognizing strings
- Algorithm for determinzing non-deterministic automata



ha!
haha!
hahaha!
hahahaha!
. . .

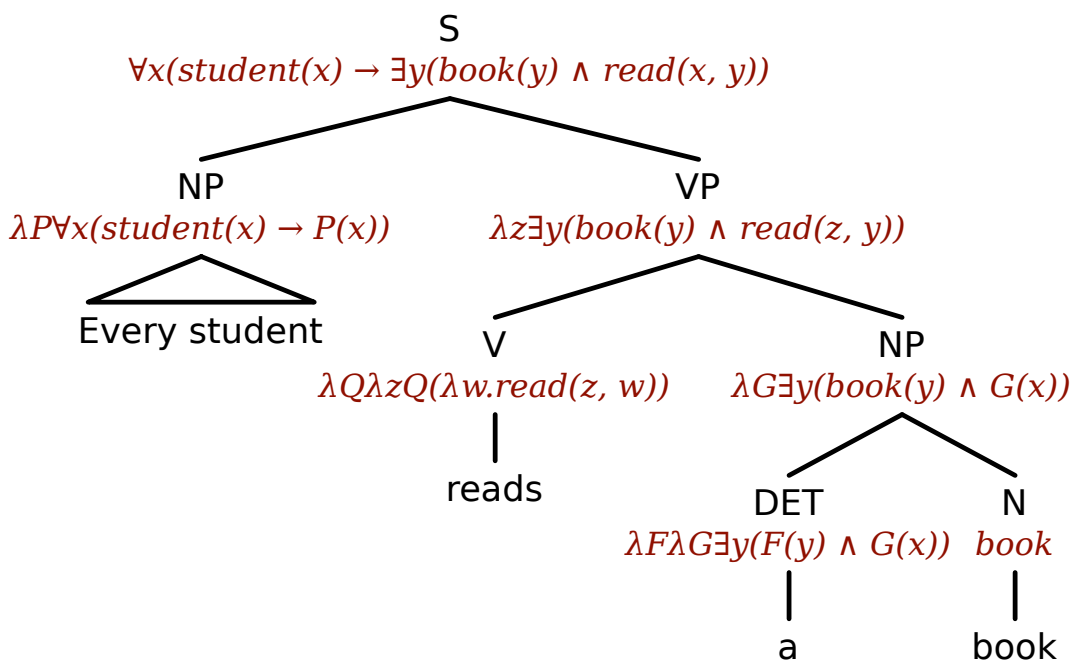
Lectures 2–4 (Thater)
Grammars & Parsing

- Context free grammars
- Bottom up parsing (shift-reduce)
- Chart parsing (CYK)
- Probabilistic grammars and parsing
- Dependency parsing

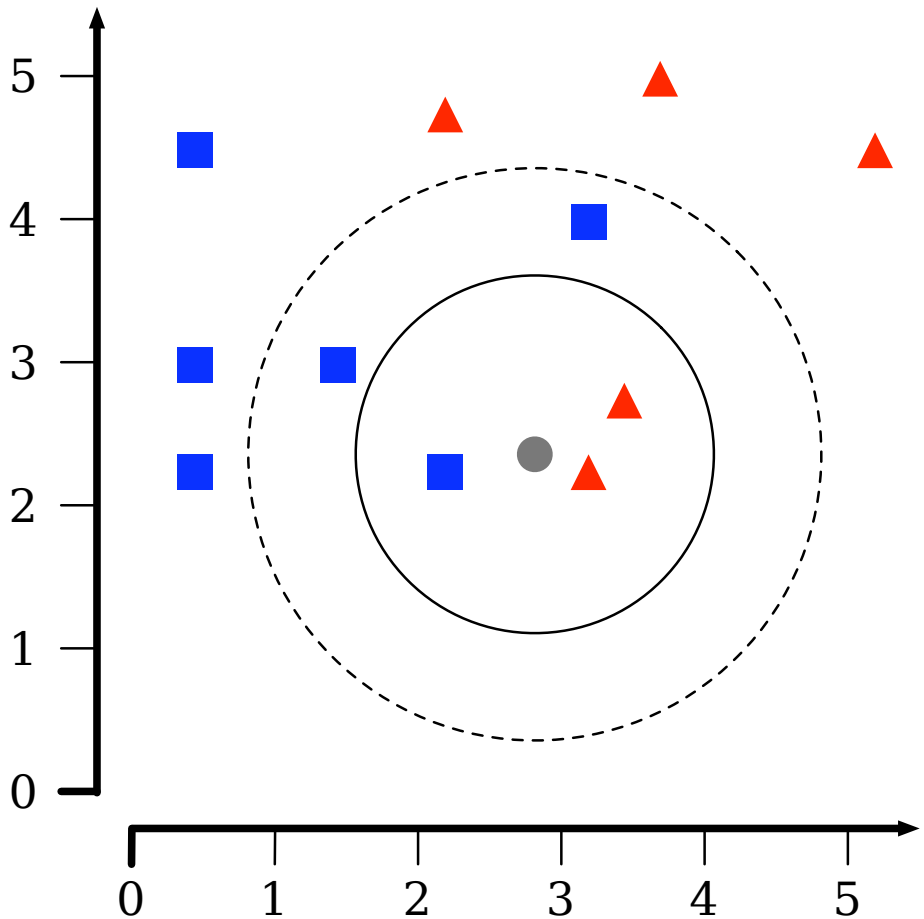


Lectures 5, 6 (Thater)
Computational Semantics

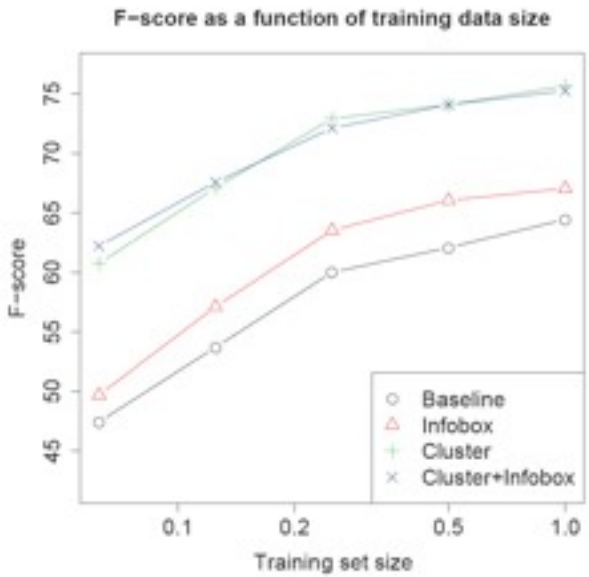
- Semantics construction
- Scope underspecification
- Dominance graphs & Tree Automata



Lecture 7 (Klakow)
Clustering



Cluster	Example members
1	Groß, Rau, Müller, Zimmermann, Frei, Becker, Möllemann, Schmidt
2	Düsseldorf, Berlin, München, Köln, Stuttgart, Hannover, Hamburg
3	nahmen, macht, zeigt, gleichen, bringt, biete, machte, sorgt, enthält



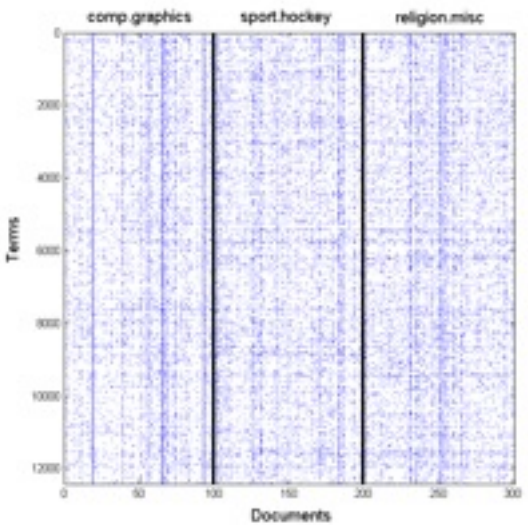
Lecture 8 (Klakow)
Matrix Factorization

- Latent semantic analysis (LSA)

- Non-negative matrix factorization (NMF)

Lecture 8 (Klakow)
Matrix Factorization

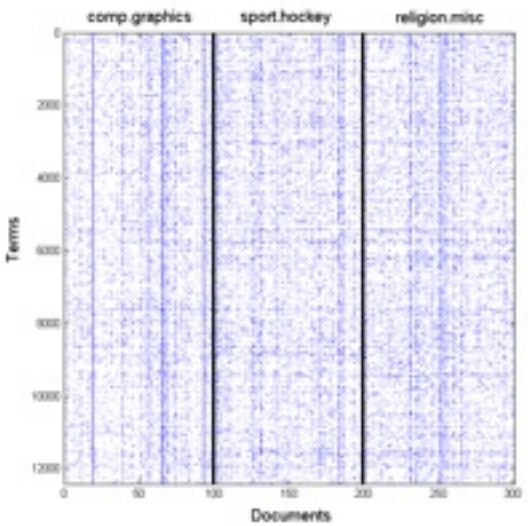
- Latent semantic analysis (LSA)

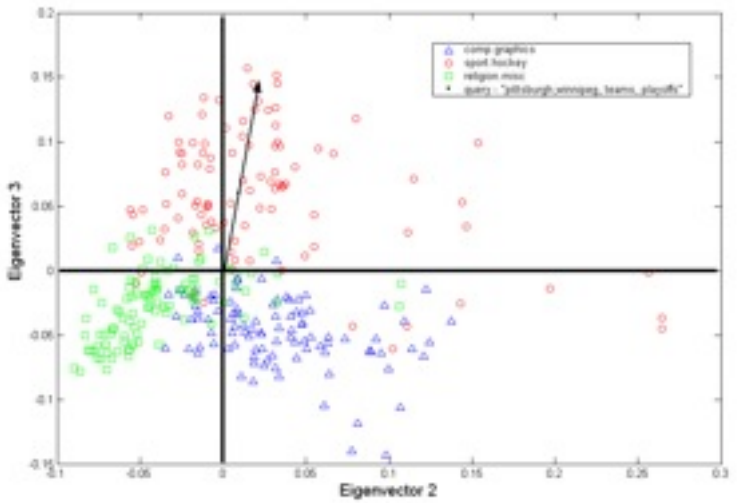


- Non-negative matrix factorization (NMF)

Lecture 8 (Klakow)
Matrix Factorization

- Latent semantic analysis (LSA)

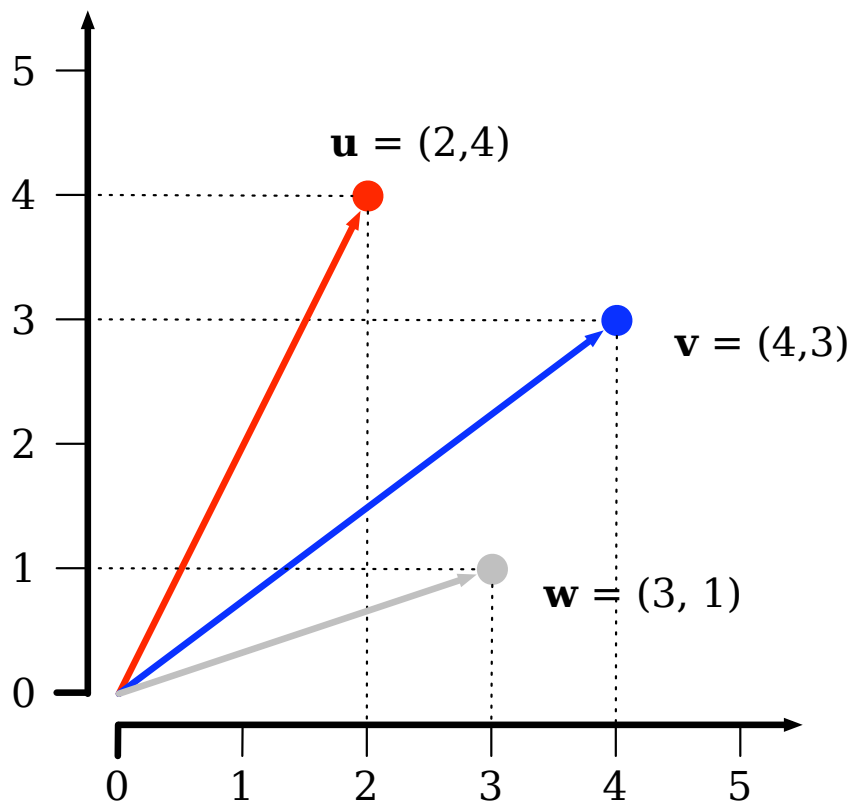




- Non-negative matrix factorization (NMF)

Lecture 9 (Klakow)
Distributional Semantics

- Distributional Semantics
- Vector space models
- Similarity measures



Lecture 10 (Klakow)
Combining Visual & Textual Data

Weather News

1 2 3 4 5

Football

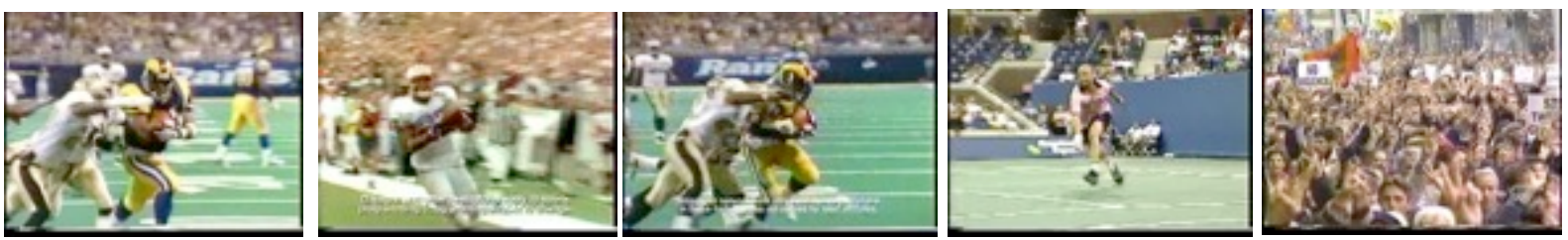
Horse

Lecture 10 (Klakow)
Combining Visual & Textual Data

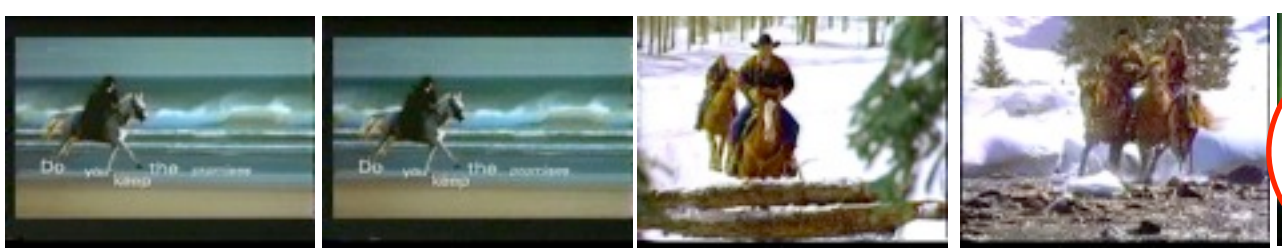
Weather News

1 2 3 4 5

Football



Horse



Alternative:
caption generation

Lecture 11 (?)
To be announced