



Automatic Text Summarization Practical Course

Automatic Text Summarization

Content and Goals

Automatic Text Summarization is the automatic generation of summaries from a collection of different types of texts

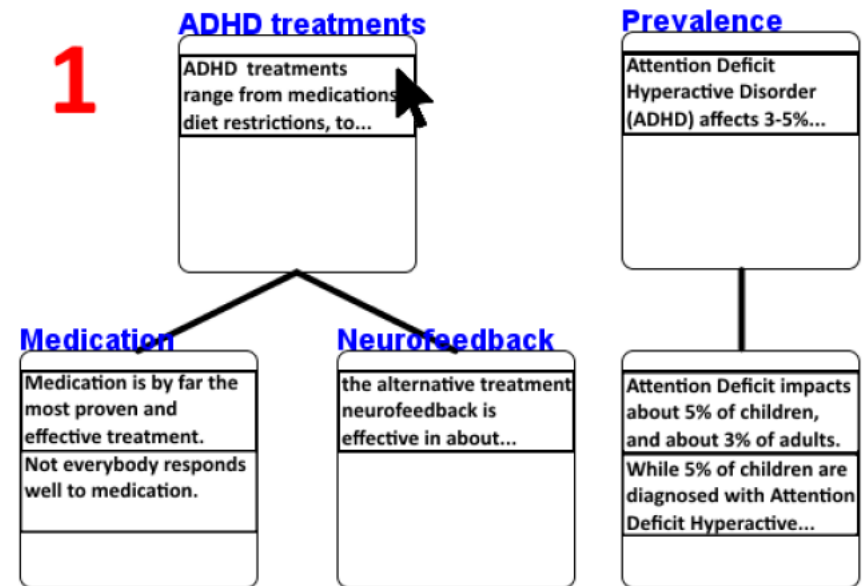
- application and development of methods
 - from machine learning, natural language processing
 - supervised, unsupervised etc.
- evaluation
 - automatic and manual assessment of the quality of resulting summaries

Automatic Text Summarization

Content and Goals

Data set

- existing corpus based on a collection of documents on topics
 - ca. 50 topics
 - 10 topics with extracted information for potential summaries, obtained via crowd sourcing and trained annotators
 - each topic contains from 31 to 90 documents
 - texts are documents of various
 - news, scientific articles, etc.
- outcome is hierarchical representation of content
 - alternatively a free text



Automatic Text Summarization

Content and Goals



Tasks

- First phase
 - application, combination, development of summarization approaches
 - goal: create summaries for each of the topics
 - hierarchical, extractive/abstractive free text ...
 - each group will submit one or several generated summaries

- Second phase
 - groups will evaluate each other's summaries
 - according to a series of evaluation criteria

Automatic Text Summarization Organization

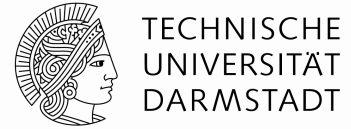
Group work

- groups of size 3-5
- will focus on particular topics
- collaboration between groups is highly encouraged

Regular meetings

- around every two weeks
- presentation and discussion of progress
- interchange between groups/groups and organizers

Automatic Text Summarization Organization



Prerequisites

- Completion of a lecture in machine learning, data mining, or NLP
- Practical experience with data mining or NLP tools is helpful, but not necessary
- Good to very good knowledge of written English
 - You will have to evaluate the quality of summaries written in English

Contact

- autots@ke.tu-darmstadt.de
- Data Sets, Text Quality Evaluation, NLP: Margot Mieskes, Thomas Arnold
- Machine Learning, Data Mining: Eneldo Loza

Automatic Text Summarization Schedule



Regular meetings

- Fridays, 13:30, Room A313

Tentative Schedule

- 13.4. Tutorial part I
- 20.4. Tutorial part II
- 4.5. Meeting
- 18.5. Meeting
- 8.6. Meeting
- 15.6. Intermediate Presentations
- 29.6. Meeting
- 13.7. Final Presentations
- 31.7. Written reports deadline

Automatic Text Summarization

Course Evaluation



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Grading

- commitment to the course
- participation in discussions/meetings
- quality of solution
 - originality of ideas
 - realization/implementation quality
 - not results (evaluation measures etc.)!
- oral presentations
 - see our general guidelines for talks
<https://www.ke.tu-darmstadt.de/lehre/arbeiten/giving-a-talk-at-a-ke-seminar-1>
- written report
- generally: groups are graded as a whole

Automatic Text Summarization Resources



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Summarization Tools

- see home page
- usually Python based, but any other tool or platform is fine

Dataset

- will be made available next week
- other datasets (e.g. for training or prototyping)

Literature

- LREC-Paper (see next week)
- see home page

Collaboration

- Moodle

Automatic Text Summarization Resources



Now: Intro to summarization!