

Open Projects

AP/FP	Topic
Initials of tutor	Given data: <Starting Point> Algorithm: <Programming Challenge> Programming: <Programming Languages> Goal: <Main Objective> Remark: <if necessary>

assigned project are grayed out

AP = Anfängerpraktikum

FP = Fortgeschrittenenpraktikum

Initials:

AS = spitz@informatik.uni-heidelberg.de

CS = sengstock@informatik.uni-heidelberg.de

JG = johanna.geiss@informatik.uni-heidelberg.de

JS = stroetgen@informatik.uni-heidelberg.de

TB = thomas.boegel@informatik.uni-heidelberg.de

FP	Temporal Tagging Clinical Data
JS	<p>Given data: Clinical TempEval Data, Thyme Corpus, HeidelTime</p> <p>Programming: Java, regular expressions</p> <p>Goal: Tuning the temporal tagger HeidelTime for the Clinical Domain, analysing the differences between clinical data and other text domains</p> <p>Remark: due to very strict licensing issues (clinical data!), the student has to work at the DBS offices</p>

FP	HeidelTime for GATE
JS	<p>Given data: HeidelTime</p> <p>Programming: Java, GATE, UIMA</p> <p>Goal: There are currently two HeidelTime versions publicly available: a UIMA version and a standalone version. The goal of this practical is to write a wrapper for GATE, i.e., that HeidelTime can be used within the GATE framework.</p>

FP	Sequence Tagging for Quotation Extraction
TB	<p>Given: Corpus with annotated quotes</p> <p>Problem: extract quotes/statements from unstructured news articles using machine learning</p> <p>Algorithm: Sequence Tagging (e.g., CRF)</p> <p>Programming: Java, Python, ML libraries</p>

FP	German Sentiment Analysis/Opinion Mining
TB	<p>Given: annotated data sets</p> <p>Problem: implementation of a sentiment analysis system for German</p> <p>Algorithm: re-implementation of lexical/statistical approaches from related work</p> <p>Programming: Python, Java</p>

FP	Tile-Raster Layer for GlobeVis (Our OpenGL Geodata Viewer)
CS	FP: Implement a Tile-Raster Layer for GlobeVis. For this you need to implement a new Layer which requests a set of geo-rectangles based on the zoom level and the camera center. The FP should provide a generic interface for arbitrary Tile-Renderers (i.e., request raster layers and show them appropriately).

FP	Implement Zoom/Rotate Animations into GlobeVis
CS	FP: Zooming and rotations should be animated. For this you need to implement a new asynchronous Animation class that updates the camera view within a certain time interval. The FP project should include a clear Animation abstraction which also works for arbitrary animations.

AP	Extension of RSS-Newsfeed extraction to British newspapers
AS	Goals: 1. Identify rules for the extraction of plain text from British online newspaper articles (rules for German newspapers exist as reference) 2. Design and implement a database connection to MongoDB for storing the extracted articles using MongoDB Java Driver. Algorithm: Existing RSS-Extraction framework (Java, uses ROME and jsoup) Programming: Java, MongoDB, basic understanding of HTML

FP	Semi-automated extraction of networks from Werewolf play-by-post games
AS	Given: URLs of internet forums for Werewolf / Mafia PbP-games Goal: Write a program that pre-processes the forum thread of a play-by-post game of Werewolf / Mafia and tags relevant actors and interactions according to a configurable set of rules. Results should be displayed in a simple, interactive GUI for manual processing and then exported to a suitable network format. Programming: Language with GUI interaction capability (your choice), basic understanding of HTML Remark: Collected Data can be used as basis for a subsequent Master's thesis if interested

FP	Person annotated Wikipedia
JG	<p>Given data: person annotated English Wikipedia (Stanford NER) Programming: Python or Java, MongoDB Goal: We have two sets of annotations for the English Wikipedia, one using interwiki-links for identifying person, the other was build using (Stanford NER). The task is to 1. remove duplicates (persons found by following interwiki-links) 2. create rules for removing obviously wrong annotation (“&quot”, “References”) 3. create basic rules for identifying and normalizing person mentions Remark: background in Computational Linguistics is an advantage</p>

AS = spitz@informatik.uni-heidelberg.de

CS = sengstock@informatik.uni-heidelberg.de

JG = johanna.geiss@informatik.uni-heidelberg.de

JS = stroetgen@informatik.uni-heidelberg.de

TB = thomas.boegel@informatik.uni-heidelberg.de

MG = gertz@informatik.uni-heidelberg.de