Maschinelle Sprachverarbeitung

Exercises - Introduction

Ulf Leser

General Structure

- We will build teams (today)
- Each team has to solve five assignments
    - Each assignment must be solved
    - We report bugs and give time for corrections
- One mandatory exercise every 2/3 weeks
    - 2-3 groups present their solution to the last assignment
    - We present new assignment
    - Each group must be present
    - Repeated misses lead to exclusion from course
- Slots in-between: Reserved for questions

Tasks and Competition

- Assignments are implementation–heavy; no paper work
  - You build tools from scratch or use existing NLP frameworks
- We provide (large) data sets to work on

- Assignments must be solved and may be solved well
  - Competition: Fastest / most accurate solution
  - Participation is voluntarily
  - Best teams get 5/3/1 points
  - Best overall team is rewarded

Assignments

- Parsing an XML corpus
- POS tagging with a Hidden-Markov-Modell
- SPAM classification
- Rule-Based NER, fuzzy dictionary matching
- NER with a CRF

Building Groups (Three each)

- Group1: Schröder, Fichte, Döpmann
- Group2: Dübel, Hoffmann, Völker
- Group3: Stiehler, Rüdian, Borchert
- Group4: Radmer, von Borries, Khanalizadeh
- Group5: Schwertfeger, Lipka, Kloke
- Group6: Ostapenko, Pahde, Demirci
- Group7: Drost, Weber, ??
- Group8: Klincov, Sterbak, Dettmer
- Group9: Krause, Goldammer, Lenk
- GroupA: Khan, Khan, Deppisch
- GroupB: Höfer, Ganschow, ???
- Please build group in GOYA as soon as possible

Questions?