



Anwendungen der KI
– Sommersemester 2018 –

Kapitel 10: Projekt-Infos

Prof. Adrian Ulges
B.Sc. {Angewandte, Medien-, Wirtschafts-}informatik
Fachbereiche DCSM
Hochschule RheinMain

Mach es zu Deinem Projekt



Vor Projektstart: Übungen komplettieren

- ▶ Blatt 04/05 (Tensorflow) ist die letzte **große** Aufgabe.
- ▶ Blatt 06/06 sind weniger umfangreich (WordNet / SPARQL).
- ▶ Wer sich ranhält, kann früher mit dem Projekt loslegen.

Projekt: Zielsetzung

- ▶ 14 Teams (*gleiche Zusammensetzung wie in Übungen*).
- ▶ 752 Trainingsfragen mit Ground Truth (*bereits online, siehe Übungsblatt 02*).
- ▶ 158 Testfragen (*noch hochgeheim!*).
- ▶ **Ihr Ziel:** Ein QA-System bauen, das (a) auf den Trainingsfragen trainiert ist, und (b) so viele **Testfragen** wie möglich richtig beantwortet.

Was dürfen wir im Projekt verwenden?



Daten

- ▶ Trainingsfragen
- ▶ Die Wikibase mit den Antworten (*siehe Übungsblatt 02*)
- ▶ Wordnet, DBPedia
- ▶ **Vortrainierte Embeddings** (*für Term-Ähnlichkeiten*)
- ▶ Weitere Daten (*nach Rücksprache*)

Software

- ▶ Programmiersprache: Python
- ▶ Sämtlicher in den Übungen entwickelter Code (Suchmaschine, Question Classifier, Synonymfinder, ...)
- ▶ Alle Bibliotheken die wir verwendet haben (*tensorflow, elasticsearch, nltk, ...*)
- ▶ **Der Stanford NER-Tagger**
- ▶ Weiteres (*nach Rücksprache*)

Gütemaß: Mean Reciprocal Rank (MRR)



- ▶ Ihr System soll für **jede Testfrage** eine **gerankte Liste** mit 10 potenziellen Antworten liefern

"Who murdered Abraham Lincoln?"

1. George R R Martin
2. Posh Spice
3. 972 km
4. **John Wilkes Booth**
- ...
10. Jack Barts

- ▶ Befindet sich die **korrekte Antwort** auf Rang 4 (*s.o.*), ist der *reziproke Rang* $r = 1/\text{Rang} = 1/4$.
- ▶ Befindet sich die korrekte Antwort **nicht** in der Liste, werten wir den reziproken Rang als 0.
- ▶ Es seien n Testfragen gegeben, und Ihr QA-System produziere zu diesen reziproke Ränge r_1, \dots, r_n . Dann ist unser **finale Gütemaß** der sogenannte **Mean Reciprocal Rank** (MRR):

$$MRR := \frac{1}{n} \sum_i r_i$$

Gütemaß: Mean Reciprocal Rank (MRR)



- ▶ Der MRR liegt immer zwischen 0 und 100%
(*je höher, desto besser*)
- ▶ Wir wollen auch nicht-exakte Treffer werten
(z.B. "John Booth" statt "John Wilkes Booth")
- ▶ Es seien T_1, T_2 die Menge der Terme der echten und der von Ihrem System ermittelten Antwort. Wir berechnen den **prozentualen Überlapp** als

$$\alpha := \frac{\#(T_1 \cap T_2)}{\#(T_1 \cup T_2)}$$

- ▶ Der reziproke Rang wird dann mit α multipliziert und somit **gewichtet**. Die **beste Antwort** Ihrer Zehnerliste geht dann in den MRR ein.

Der Preis...



- ▶ Der beste Score des letzten Jahres: 33%.
- ▶ Das beste Team erhält einen **PREIS** (tbd) !



Anwendungen der Künstlichen Intelligenz (SS17)

Scoreboard Question Answering

High Scores

rank	team	score(percent)
1.	Artificial Stupidity	33.0
2.	Team Triggered	30.0
3.	Die eisernen Einhoerner	28.3



XX.XX.2018: Finaltag

- ▶ Wahrscheinlich Ende Juli / Anfang August.
- ▶ Testdaten werden morgens online gestellt.
- ▶ Abgabe der Lösung abends (csv-Format).
- ▶ **Nur *vollautomatische* Lösungen, ehrlich bleiben!**
- ▶ Benotet wird **nicht nach Score**.
- ▶ Scoreboard wird online gestellt.

Fachgespräche

- ▶ 08.-10. August
- ▶ Jedes Team: Vortrag mit Slides
- ▶ Anschließendes Gespräch zu Projekt und theoretischen Grundlagen (Vorlesung)



1. Iterativ vorgehen!

- ▶ Bauen Sie am Anfang ein **sehr einfaches** System
- ▶ Entwickeln Sie Ihr System **iterativ** weiter:
Experimentieren → Messen → System ändern/verbessern
- ▶ Analysieren Sie Fehler: Was ist der “Flaschenhals”?

2. Dokumentieren!

- ▶ Führen Sie ein regelmäßiges **Lab Notebook**
 - ▶ Was genau wurde ausprobiert?
 - ▶ Welche Parameter wurden wie gewählt?
 - ▶ Was wurde gemessen / beobachtet?
 - ▶ Ziel für nächste Woche?
- ▶ Dokumentiert den **Projektverlauf**
- ▶ Grundlage für wöchentliche Diskussionen



3. Kreativ sein!

- ▶ Für einige Teilprobleme haben wir fertige Lösungen aus den Übungen (IR, Question Classification...).
- ▶ Andere Teilprobleme sind 'offen' zu lösen (*insbesondere: Antwort-Auswahl*).
- ▶ Welche **Datenquellen** / **Ansätze** könnten für die Lösung interessant sein?
- ▶ Probieren Sie ruhig ungewohnte Sachen aus, diskutieren Sie mit Ihrem Lieblingsprof!

4. Rechenpower benötigt?

- ▶ per E-Mail anfragen!