DIS22:
2. Meeting: Besprechung des Papers, Aufgabenvorstellung

Philipp Schaer, Björn Engelmann, Fabian Haak

CIR
Technology
Arts Sciences
TH Köln

Methodology

ure 1 gives an overview of our approach. We
select and prepare texts and match them into
s. These pairs are rated and an Elo algorithm is
lied to the rated pairs. Based on that, the ARTS
plicity ranking is composed.

Rationale b[illegible] System

bdir et al. (2023) describe that the Elo rating
em was increasingly used to compare Large
guage Models (LLMs) through "A vs. B" paired
nparisons. We build on this idea and evaluate
s instead of LLMs to assign them a score in-
ating their simplicity. To assess the simplicity

matches, the expected probability E_{T_1} that T_1 w
Elo rating R_1 wins against T_2 with Elo rating
is defined as follows (Good, 1955; Boubdir et
2023):

$$E_{T_1} = \frac{1}{1 + 10^{(R_2 - R_1)/400}}.$$

After a victory of T_1 over T_2 ($T_1 > T_2$), the n
rating [illegible] is calculated as follows :

$$R'_1 = R_1 + k \cdot E_{T_1}.$$

The constant k controls how strong the chan
should be after a game. In our case, k is set to
The expected probability of winning for T_2 and
calculation of the score update are calculated ar

1. Hausaufgabe: ARTS lesen

2. Vorstellung der Aufgaben

Philipp Schaer, Björn Engelmann, Fabian Haak

Technology
Arts Sciences
TH Köln

Zwischenfragen: Wissensabfrage

Ziel ist es, Zwischenfragen zu implementieren, die Inhalte zu vorherigen Texten abfragen. Damit soll überprüft werden können, ob die Komplexität von Texten damit korreliert, wie sehr Inhalte (langfristig) vom Leser memoriert werden.

- Einlesen ins Thema
  - Relevante Quellen finden: Wissenschaftliche Auseinandersetzung damit, wie sehr bestimmte Arten von Fragemodalitäten (z.B. Lückentext, etc.) geeignet sind, das Textverständnis zu überprüfen.
- Abfrageschema implementieren
  - Fragen z.B. nach Multiple Choice, Lückentext, Wahr/Falschaussagen, … oder Kombinationen!
- Fragen zu Texten entwickeln
  - Dies kann, je nach Abfrageschema, manuell oder automatisiert, z.B. über ChatGPT erfolgen. Es sollte am Ende mindestens 50 Fragen formuliert sein.
- Implementierung in Streamlit Interface
  - Sollte immer nach X Vergleichen oder in zufälligen Intervallen gefragt werden? Sollte direkt zu den letzten Texten gefragt werden? Sollte nur zu Texten gefragt werden, die der Nutzer zum ersten mal gelesen hat?

2 Personen

Qua L itativ

C oding

Philipp Schaer, Björn Engelmann, Fabian Haak

Technology
Arts Sciences
TH Köln

Kenntnisse zu Themen abfragen

Es soll erhoben werden, wie vertraut die Personen mit den Themen der Texte sind.
Dazu müssen zunächste über Clustering (sprecht mit Björn und Fabian für Tipps), mit ChatGPT oder manuell einer Reihe von Themen zugeordnet werden. Dann soll ein System implementiert werden, mit dem die Kenntnisstände der Nutzer erfasst werden, z.B. eine Reihenfolge erstellen lassen.

2 Personen

- Einlesen ins Thema
- Texte nach Themen klassifizieren
- Kenntnisstanderhebung implementieren
- Implementierung in Streamlit Interface

Qua **L** itativ

C oding

Philipp Schaer, Björn Engelmann, Fabian Haak

Technology
Arts Sciences
TH Köln

Konfidenzabfrage, Nutzerführung und Nutzerprofil

Es gilt zu messen, wie sicher sich eine Person über ihr Urteil ist (Likert-Skala). Dies ist so zu implementieren, dass möglichst wenig Aufwand für die Personen entsteht. Zusätzlich soll für eine bessere Nutzerführung eine Fortschrittsanzeige und ein Nutzerprofil eingeführt werden. Schöner machen ist auch immer gut.

1 Person

Coding

- Likert-Skala festlegen
- Fortschrittsanzeige in Prozent
- Nutzerprofil mit Login
  - Username (Pseudonym)
  - Englisch Level
  - Alter
  - Rating History
  - ...

Fragenreihenfolge

Diese Aufgabe beinhaltet etwas mehr wissenschaftliche Konzeptionsarbeit.
Zu den Texten werden wir ein automatisch erzeugtes GPT-Rating zur Verfügung stellen. Ziel soll es sein, die Nutzer (geheim) zufällig bestimmten Fragereihenfolgen zuzuordnen, die dann unterschiedliche Paarzusammenstellungen und -reihenfolgen erhalten. Für eine wiss. Auswertbarkeit erzähl bitte nur den Betreuern von den Reihenfolgen, die ihr konzipiert!

1 Person

Konzeption

Coding

- Reihenfolgen konzipieren
  (z.B. zuerst Paare mit geringer Differenz, dann immer “deutlichere” Paare und umgekehrt)
- Paare im Code implementieren
- Zuordnung zu Benutzern implementieren

Philipp Schaer, Björn Engelmann, Fabian Haak

Technology
Arts Sciences
TH Köln

Logger

1 Person

Da das Logging für die spätere Auswertung essentiell ist, soll sich eine Person in enger Zusammenarbeit mit den anderen Gruppen um das Logging kümmern.

Konzeption

Coding

Philipp Schaer, Björn Engelmann, Fabian Haak

Technology
Arts Sciences
TH Köln

Bonus: Eure Tasks!

Hier sind Eure Ideen gefragt! Was fällt euch noch ein?

Interrater Agreement

- Vergleich verschiedener Vergleichswerte wie Fleiss/Cohen Kappa, Krippendorf Alpha, etc.
- Experimente mit verschiedenen Schwellwerten und dem Einfluss auf die “Güte” der Bewertungen,
  Beispiel:
  https://arxiv.org/ftp/arxiv/papers/1206/1206.4802.pdf

Qua **N** titativ

K onzeption

Biases

- Werden Texte die als erstes/zweites genannt werden tendenziell simpler bewertet?
- Werden Texte die man schon häufiger gesehen hat tendenziell als einfacher bewertet?
- Gibt es andere Faktoren, die die Ergebnisse beeinflussen?

Antwortdauer und Simplizität

- Korrelation zwischen Vertrautheit mit Texten (wie oft gesehen) (Lerneffekt)
- Korrelation mit Anzahl beantwortetet Fragen (Lerneffekt)
- Korrelation mit Textkomplexität
- Korrelation mit Textlänge
- Korrelation mit …

2 Personen

Qua N titativ

Biases

- Werden Texte die als erstes/zweites genannt werden tendenziell simpler bewertet?
- Werden Texte die man schon häufiger gesehen hat tendenziell als einfacher bewertet?
- Gibt es andere Faktoren, die die Ergebnisse beeinflussen?

Bonus: Eure Tasks!

- Memoriertheitsabfragen
- Themen Kenntnisstand
- Fragenreihenfolge
- …

Datum	Dauer	Was steht an?
11.04.2024	90 min	Kick-Off, Vorstellung der Gruppenaufgaben
18.04.2024	90 min	Einteilung Aufgaben, Besprechung ARTS Paper
25.04.2024	180 min*	**Vorstellung Zeitplan + Labeling Session**
02.05.2024	90 min	**Deadline Labeling Session**
09.05.2024		CHRISTI HIMMELFAHRT
16.05.2024	flexibel	Projektwoche (bei Bedarf: Implementationsgruppe)
23.05.2024	90 min	Flexible Meetingslots
30.05.2024		FRONLEICHNAM
06.06.2024	90 min	Vorstellung/Abnahme Implementation
13.06.2024	180 min*	**Präsentation Implementationsgruppe, Labeling Session 2**
20.06.2024	90 min	**Deadline Labeling Session 2**
27.06.2024	90 min	Flexible Meetingslots
04.07.2024	90 min	Flexible Meetingslots
11.07.2024	90 min	**Abschlusspräsentation**