

Introduction

HS Rankingaufgaben in der Computerlinguistik

Nils Reiter

`nils.reiter@uni-koeln.de`

Department of Digital Humanities

April 6, 2023

(Sommersemester 2023)

HS Rankingaufgaben in der Computerlinguistik

- ▶ „Rankingaufgaben“ / Ranking tasks
 - ▶ Put things in some order
 - ▶ Contrary to classification or sequence labeling, in which we want to assign categories to things
 - ▶ Easiest example: Rank documents according to some search query
- ▶ „Computerlinguistik“ / Computational linguistics
 - ▶ Discipline dealing with processing language with a computer
 - ▶ Related to natural language processing, digital humanities

Wer sind Sie?



Vorstellungsrunde!

- ▶ Wer sind Sie?
- ▶ Warum sind Sie hier?
 - ▶ Hier: in diesem Kurs, in diesem Modul, in dieser Studienrichtung, in diesem Studiengang, ...?
- ▶ Was wissen Sie über Computerlinguistik?
- ▶ Wo wären Sie jetzt, wenn Sie sich mal anders entschieden hätten?
 - ▶ Z.B. bei der Feuerwehr?

Lernziele

- ▶ Lesen und Verstehen von (computerlinguistischer) Forschungsliteratur
- ▶ Aufbereitung eines Themas für andere
- ▶ Überblick über Ranking-Verfahren und ihre Anwendungsmöglichkeiten

Section 1

Organisatorisches

Module und Studiengänge

- ▶ B.A. Informationsverarbeitung: Angewandte linguistische Datenverarbeitung / Anwendungen der Computerlinguistik
- ▶ M.A. Informationsverarbeitung: Verarbeitung von Textdaten
- ▶ M.A. Informatik: Aufbaubereich Digital Humanities

Module und Studiengänge

- ▶ B.A. Informationsverarbeitung: Angewandte linguistische Datenverarbeitung / Anwendungen der Computerlinguistik
- ▶ M.A. Informationsverarbeitung: Verarbeitung von Textdaten
- ▶ M.A. Informatik: Aufbaubereich Digital Humanities
- ▶ Alle
 - ▶ Kontaktzeit 30h / Selbststudium 60h (Informatik: 240h, aber inklusive Modulprüfung)

Ablauf des Seminars

- ▶ Drei Teile
 - ▶ Meta-Themen
 - ▶ Allgemeiner methodischer Input
 - ▶ Praktische Anwendungen in studentischen Sitzungen

Studentische Sitzungen

Ziel: Inhalt vermitteln, nicht: Papier referieren

- ▶ Mehr als zwei Wochen vorher: Literatur durcharbeiten, inklusive Forschungsstand
- ▶ Spätestens zwei Wochen vorher: Sprechstunde mit Dozent zu inhaltlichen Fragen zum Papier
- ▶ Spätestens eine Woche vorher: Konzept zum Referat per E-Mail abgeben. Das Konzept beinhaltet
 - ▶ die Gliederung/Struktur,
 - ▶ Aufgaben/Fragestellungen für Kleingruppen sowie erwartete Ergebnisse und
 - ▶ selbst erstellte Beispiele, die das gesagte verdeutlichen und konkretisieren
- ▶ Referat halten. Jedes Referat muss mind. eine Kleingruppenaktivität beinhalten
- ▶ Im Anschluss: Feedbackgespräch

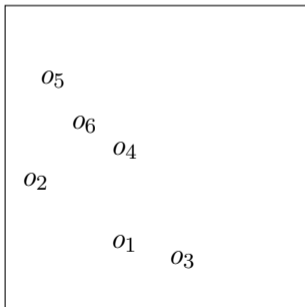
Which ranking problems do you know?

Application Areas

- ▶ May 25
 - ▶ Document Ranking, based on a search query
- ▶ June 15
 - ▶ Paraphrasing: Rewrite a statement
- ▶ June 22
 - ▶ Summarization
- ▶ June 29
 - ▶ Authorship attribution
- ▶ July 6
 - ▶ Language similarity and identification

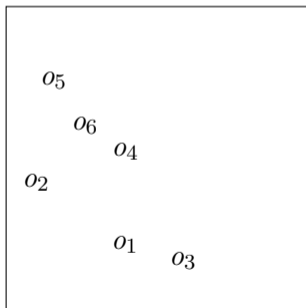
Please select three wishes in Ilias!

Terminology (Li, 2014)



Offerings $O = \{o_1, o_2, \dots, o_n\}$

Terminology (Li, 2014)

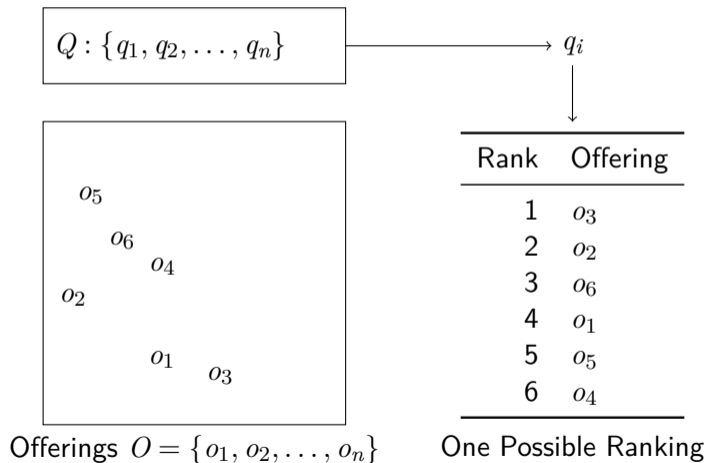


Offerings $O = \{o_1, o_2, \dots, o_n\}$

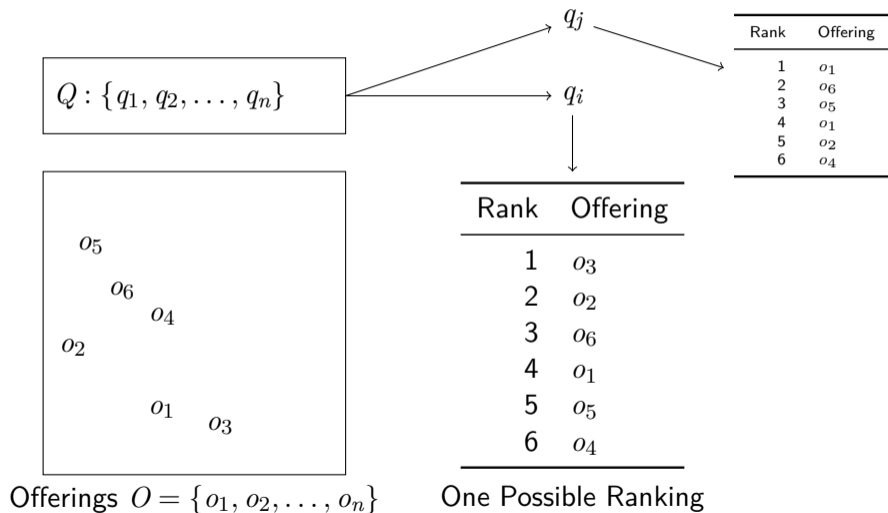
| Rank | Offering |
|------|----------|
| 1 | o_3 |
| 2 | o_2 |
| 3 | o_6 |
| 4 | o_1 |
| 5 | o_5 |
| 6 | o_4 |

One Possible Ranking

Terminology (Li, 2014)



Terminology (Li, 2014)



Schema (Li, 2014)

- 1 Offerings $O = \{o_1, o_2, \dots, o_n\}$, requests $Q = \{q_1, q_2, \dots, q_n\}$
- 2 Given one request q_i

Schema (Li, 2014)

- 1 Offerings $O = \{o_1, o_2, \dots, o_n\}$, requests $Q = \{q_1, q_2, \dots, q_n\}$
- 2 Given one request q_i
- 3 $F(q_i, O)$: Assign each $o \in O$ a score based on q_i
- 4 Sort according to score

Schema (Li, 2014)

- 1 Offerings $O = \{o_1, o_2, \dots, o_n\}$, requests $Q = \{q_1, q_2, \dots, q_n\}$
- 2 Given one request q_i
- 3 Select a subset with candidates O' from O : $O' \subseteq O$
- 4 $F(q_i, O')$: Assign each $o \in O'$ a score based on q_i
- 5 Sort according to score

Schema (Li, 2014)

- 1 Offerings $O = \{o_1, o_2, \dots, o_n\}$, requests $Q = \{q_1, q_2, \dots, q_n\}$
- 2 Given one request q_i
- 3 Select a subset with candidates O' from O : $O' \subseteq O$
- 4 $F(q_i, O')$: Assign each $o \in O'$ a score based on q_i
- 5 Sort according to score

⚠ Function F looks at entire set of candidates (Global Ranking Function)

Schema (Li, 2014)

- 1 Offerings $O = \{o_1, o_2, \dots, o_n\}$, requests $Q = \{q_1, q_2, \dots, q_n\}$
- 2 Given one request q_i
- 3 Select a subset with candidates O' from O : $O' \subseteq O$
- 4 $f(q_i, o_i)$: Assign each $o \in O'$ a score based on q_i , but separately
- 5 Sort according to score

⚠ Function f looks at one candidate at a time (Local Ranking Function)

Learning to Rank

- ▶ Field of machine learning
- ▶ Focus on predicting rankings
- ▶ Learn $F(q_i, O)$ and/or $f(q_i, o_j)$ from data

References I



Li, Hang (2014). *Learning to Rank for Information Retrieval and Natural Language Processing*. Ed. by Graeme Hirst. 2nd ed. Synthesis Lectures on Human Language Technologies. Morgan & Claypool.