

Modulbeschreibung

39-Inf-IR Information Retrieval

Technische Fakultät

Version vom 31.01.2026

Dieses Modulhandbuch gibt den derzeitigen Stand wieder und kann Änderungen unterliegen. Aktuelle Informationen und den jeweils letzten Stand dieses Dokuments finden Sie im Internet über die Seite

<https://ekvv.uni-bielefeld.de/sinfo/publ/modul/26787746>

Die jeweils aktuellen und gültigen Regelungen im Modulhandbuch sind verbindlich und konkretisieren die im Verkündungsblatt der Universität Bielefeld veröffentlichten Fächerspezifischen Bestimmungen.

39-Inf-IR Information Retrieval

Fakultät

Technische Fakultät

Modulverantwortliche*r

Prof. Dr. Philipp Cimiano

Turnus (Beginn)

Wird nicht mehr angeboten

Leistungspunkte

10 Leistungspunkte

Kompetenzen

Nach Abschluss des Moduls sollten die Studierenden:

- mit den wesentlichen Paradigmen sowie den gängigen Methoden und Modellen des Information Retrievals vertraut sein,
- Techniken beherrschen, um große Mengen an unstrukturierten Daten im Hinblick auf das effiziente Retrieval zu indizieren,
- ein Verständnis für fortgeschrittene Techniken wie die Verwendung von Sprachmodellen für das IR, relevance feedback sowie latent semantic indexing entwickelt haben,
- die Funktionsweise von Web-Suchmaschinen verstehen, sowie
- in der Lage sein, ein IR System selbstständig zu implementieren.

After completion of the module students should

- be familiar with the basic paradigms as well as the current methods and models of information retrieval
- know the techniques necessary for indexing large amounts of unstructured data for efficient retrieval
- have developed an understanding of advanced techniques like the application of language models for IR, relevance feedback as well as latent semantic indexing
- understand the functionality of search engines on the web
- be capable to implement an IR system without assistance

Lehrinhalte

Das Auffinden relevanter Informationen stellt eine zentrale Aktivität in unserer modernen Wissensgesellschaft dar. Relevante Informationen sind zum großen Teil in unstrukturierten Dokumenten (insbesondere Textdokumente) zu finden. Das Gebiet des Information Retrieval (IR) beschäftigt sich mit der Erforschung, Entwicklung und Anwendung von Methoden für den effizienten Zugriff und Suche auf großen Mengen von unstrukturierten Daten, insbesondere Texte, Bilder und Videos. In der Vorlesung werden folgende Themen behandelt:

- Information Retrieval Modelle (insbesondere das Boolesche, das vektor-basierte und das probabilistische Modell)
- Methoden zur Gewichtung von Termen
- Techniken zur Indizierung
- Sprachmodelle für das Information Retrieval
- Relevance Feedback und Query Expansion für das Information Retrieval
- Latent Semantic Indexing
- Web Suche: Der Fall Google
- Multimedia Retrieval

Ziel des Praktikums ist es, eine eigene Suchmaschine für einen größeren Datensatz (den Reuters Datensatz) zu entwickeln.

Literatur:

- "Introduction to Information Retrieval", Manning, Raghavan, Schütze, Cambridge University Press, 2008
- "Search Engines: Information Retrieval in Practice", Bruce Croft, Donald Metzler, Trevor Strohman, Pearson /Addison-Wesley, 2009
- "Modern Information Retrieval", Baeza-Yates and Ribeiro-Neto, Addison-Wesley, 1999

The retrieval of relevant information constitutes a central activity in our modern knowledge-based society. Relevant information is to a large extent found in unstructured documents (especially text documents). The area of Information Retrieval (IR) is concerned with the research, development and application of methods that allow large amounts of unstructured data, especially texts, pictures and videos, to be accessed and searched efficiently. The following topics will be discussed:

- IR models (including Boolean, vector based and probabilistic)
- term weighting methods
- indexing techniques
- language models for IR
- relevance feedback and query expansion for IR
- latent semantic indexing
- Web search: the case of google
- Multimedia retrieval

Aim of the practical is the development of a search engine for a dataset.

Empfohlene Vorkenntnisse

Algorithmen und Datenstrukturen, Grundkenntnisse Mathematik, Einführung in die Datenbanken und Modellierung

Algorithms and datastructure, fundamentals in mathematics, introduction to databases and modelling

Notwendige Voraussetzungen

—

Erläuterung zu den Modulelementen

Die Modul(teil)prüfung kann in einigen Studiengängen nach Wahl der Studierenden auch "unbenotet" erbracht werden. Vor Erbringung ist eine entsprechende Festlegung vorzunehmen, eine nachträgliche Änderung (benotet - unbenotet) ist ausgeschlossen. Wird diese Option gewählt, ist es nicht möglich, dieses Modul zu verwenden, um es in einen Studiengang einzubringen, in dem dieses Modul bei der Gesamtnotenberechnung berücksichtigt wird.

The (partial) examination of the module can be performed as "ungraded" in some study programs at the students choice. Before the examination a respective determination must be carried out, a later modification (graded - ungraded) is impossible. If the "ungraded" option is chosen, it is not possible to include this module in a study program where this module is deemed to enter the calculation of the overall grade.

Modulstruktur: 0-1 bPr, 0-1 uPr ¹

Veranstaltungen

Titel	Art	Turnus ⁵	Workload	LP ²
Einführung in das Information Retrieval	Vorlesung	WiSe	60 h (30 + 30)	2 [Pr] [Pr]
Einführung in das Information Retrieval	Übung	WiSe	60 h (30 + 30)	2
Praktikum Information Retrieval	Praktikum	SoSe	120 h (60 + 60)	4

Prüfungen

Zuordnung Prüfende	Art	Gewichtung	Workload	LP ²
<p>Lehrende der Veranstaltung Einführung in das Information Retrieval (Vorlesung)</p> <p><i>In einigen Studiengängen der Technischen Fakultät kann die Modulprüfung nach Wahl der Studierenden auch "unbenotet" erbracht werden (s. Erläuterungen zu den Modulelementen und die jeweilige FsB). Wird die unbenotete Option gewählt, ist es nicht möglich, dieses Modul zu verwenden, um es in einen Studiengang einzubringen, in dem dieses Modul bei der Gesamtnotenberechnung berücksichtigt wird. Erläuterungen zu dieser Prüfung siehe unten (benotete Prüfungsvariante).</i></p>	Portfolio mit Abschlussprüfung	unbenotet	60h	2

<p>Lehrende der Veranstaltung Einführung in das Information Retrieval (Vorlesung)</p> <p><i>Portfolio aus Übungsaufgaben, die veranstaltungsbezogen gestellt werden (Bestehensgrenze 60% der erzielbaren Punkte, individuelles Erläutern von Aufgaben) und Abschlussklausur (60-90 min). Die Übungsaufgaben im Rahmen des Portfolios werden in der Regel wöchentlich ausgegeben.</i></p> <p><i>Abschlussklausur bezieht sich auf den Stoff der Vorlesung und der Übungen.</i></p> <p><i>Exercises need to be completed on a weekly basis covering material from the lectures (pass mark of 60%, students have to be able to demonstrate how they arrived at the solution). Final written examination (60-90min) relating to material covered in both lectures and tutorials.</i></p>	Portfolio mit Abschlussprüfung	1	60h	2
---	--------------------------------	---	-----	---

Weitere Hinweise

Bei diesem Modul handelt es sich um ein eingestelltes Angebot. Ein entsprechendes Angebot, um dieses Modul abzuschließen, wurde bis maximal Sommersemester 2018 vorgehalten.
Bisheriger Angebotsturnus war jedes Wintersemester.

Legende

- 1 Die Modulstruktur beschreibt die zur Erbringung des Moduls notwendigen Prüfungen und Studienleistungen.
 - 2 LP ist die Abkürzung für Leistungspunkte.
 - 3 Die Zahlen in dieser Spalte sind die Fachsemester, in denen der Beginn des Moduls empfohlen wird. Je nach individueller Studienplanung sind gänzlich andere Studienverläufe möglich und sinnvoll.
 - 4 Erläuterungen zur Bindung: "Pflicht" bedeutet: Dieses Modul muss im Laufe des Studiums verpflichtend absolviert werden; "Wahlpflicht" bedeutet: Dieses Modul gehört einer Anzahl von Modulen an, aus denen unter bestimmten Bedingungen ausgewählt werden kann. Genauereres regeln die "Fächerspezifischen Bestimmungen" (siehe Navigation).
 - 5 Workload (Kontaktzeit + Selbststudium)
- SoSe** Sommersemester
- WiSe** Wintersemester
- SL** Studienleistung
- Pr** Prüfung
- bPr** Anzahl benotete Modul(teil)prüfungen
- uPr** Anzahl unbenotete Modul(teil)prüfungen