

Hausarbeit , Richtwert 25000 Zeichen fortlaufender Text. Abgabe bis spätestens 25. August ausgedruckt im Sekretariat des Centrum für Informations und Sprachverarbeitung und in elektronischer Form PDF unverschlüsselt beim Dozenten per email an [kristof@cis.uni-muenchen.de](mailto:kristof@cis.uni-muenchen.de). Aufgrund des Notenschlusses wird es keine Nachfrist für die Abgabe geben.

Hinweis zur Zusammenarbeit: Sie sollen und dürfen die Literatur zusammen erarbeiten. Dann allerdings sollte Ihre Zusammenarbeit ein Ende haben. Die Arbeiten werden vor der Bewertung durch ein Plagiatserkennungsprogramm laufen.

## Hausarbeit Informatik 6 Ects

Beschreiben Sie die Elemente eines Informationretrievalsystems wie es im Begleitbuch zur Vorlesung Manning /Prabhakar/ Schütze dargestellt wird. Erklären Sie das Vektorraummodell mit seinen zentralen Komponenten. Erklären Sie Unterschiede, die für Retrieval auf strukturierten Dokumenten wichtig sind und Unterschiede für das Webretrieval. Was ist die Idee des probabilistisch basierten IR, machen Sie die Unterschiede zum Standard Vektorraummodell klar. Ergänzen Sie Ihre Ausführungen indem Sie an geeigneten Stellen die nachfolgend angegebene Literatur einarbeiten. Wo etwa geht das Papier Zobel und Moffat über das in Vorlesung und Buch angebotene hinaus, wie wirken sich bestimmte Indexarchitekturen auf die Abarbeitung operatorenverbundener Queries aus. Was ist Clickthrough-Analyse und wie könnte sie in das probabilistische Modell des Information Retrieval eingebracht werden. Gehen Sie schließlich auf die Websuche ein. Was war das revolutionär neue an dem Modell von Brin und Page und welche Schwierigkeiten gibt es für neue Suchmaschinen dieses Modell nachzuvollziehen. Ein Hinweis: versuchen Sie die Zusatzliteratur quer zu lesen. Im Gegensatz zu den formalen Stellen im Lehrbuch ist es nicht wichtig jedes Detail der mathematischen Darlegung zu verstehen sondern die zentralen Aussagen und Ihre Wichtigkeit für die Verbesserung des Retrievals zu bewerten und in Ihren Text zu integrieren.

## Indexierung

Zobel, Justin, and Alistair Moffat. 2006. Inverted files for text search engines. *ACM Computing Surveys* 38(2). 18, 83, 106, 133, 528, 533

## Clickthrough

Joachims, Thorsten. 2002b. Optimizing search engines using clickthrough data. In *Proc. KDD*, pp. 133–142. 175, 185, 348, 525

Joachims, Thorsten, Laura Granka, Bing Pan, HeleneHembrooke, and Geri Gay. 2005. Accurately interpreting clickthrough data as implicit feedback. In *Proc. SIGIR*, pp. 154–161. ACM Press. 175, 185, 523, 524, 525, 528

## Websuche

Brin, Sergey, and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. In *Proc. WWW*, pp. 107–117. 149, 458, 480, 520, 528

## Hausarbeit Computerlinguistik und Informatik 9 Ects

Lesen Sie zusätzlich zu den oben angegebenen Papieren Hollink et al. und gehen Sie zusätzlich zur oben gestellten Aufgabe genauer auf die möglichen Verbesserungen eines IR Systems durch linguistische Vorverarbeitung ein.

Hollink, Vera, Jaap Kamps, Christof Monz, and Maarten de Rijke. 2004. Monolingual document retrieval for European languages. IR 7(1):33–52. 46, 524, 525, 528, 529