

Konzeption und Realisierung eines Systems zur Informationssuche in einem Dokumentenarchiv basierend auf Textinhalt und Metadaten.

Conception and Realization of an Information Retrieval System for a Document Archive based on Text Content and Metadata

Annika Kremer

Bachelor-Abschlussarbeit

Betreuer: Prof. Dr. Karl Hans Bläsius

Trier, Abgabedatum

---

## **Vorwort**

Ein Vorwort ist nicht unbedingt nötig. Falls Sie ein Vorwort schreiben, so ist dies der Platz, um z.B. die Firma vorzustellen, in der diese Arbeit entstanden ist, oder einigen Leuten zu danken, die in irgendeiner Form positiv zur Entstehung dieser Arbeit beigetragen haben. Auf keinen Fall sollten Sie im Vorwort die Aufgabenstellung näher erläutern oder vertieft auf technische Sachverhalte eingehen.

---

## Kurzfassung

In der Kurzfassung soll in kurzer und prägnanter Weise der wesentliche Inhalt der Arbeit beschrieben werden. Dazu zählen vor allem eine kurze Aufgabenbeschreibung, der Lösungsansatz sowie die wesentlichen Ergebnisse der Arbeit. Ein häufiger Fehler für die Kurzfassung ist, dass lediglich die Aufgabenbeschreibung (d.h. das Problem) in Kurzform vorgelegt wird. Die Kurzfassung soll aber die gesamte Arbeit widerspiegeln. Deshalb sind vor allem die erzielten Ergebnisse darzustellen. Die Kurzfassung soll etwa eine halbe bis ganze DIN-A4-Seite umfassen.

Hinweis: Schreiben Sie die Kurzfassung am Ende der Arbeit, denn eventuell ist Ihnen beim Schreiben erst vollends klar geworden, was das Wesentliche der Arbeit ist bzw. welche Schwerpunkte Sie bei der Arbeit gesetzt haben. Andernfalls laufen Sie Gefahr, dass die Kurzfassung nicht zum Rest der Arbeit passt.

The same in english.

---

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung und Problemstellung</b>	<b>1</b>
1.1	Einleitung	1
1.2	Problemstellung	1
1.3	Teilprobleme	2
1.3.1	Dynamisches einlesen der Metadaten	2
1.3.2	Unterscheidung Metadaten und Freitext	2
1.3.3	Metadatensuche	2
1.3.4	Freitextsuche	2
1.3.5	Verknüpfung mit UND/ODER	3
1.3.6	Benutzeroberfläche	3
<b>2</b>	<b>Information Retrieval</b>	<b>4</b>
2.1	Bedeutung	4
2.1.1	Information	4
2.1.2	Information Retrieval	5
2.1.3	Unterschied zur Datenbankensuche	5
2.2	Beispiel Websuche	5
2.3	Bezug zur Problemstellung	6
2.3.1	Teilweise strukturierte Daten	6
<b>3</b>	<b>Boolesches Retrieval</b>	<b>7</b>
3.1	Eigenschaften des Verfahrens	7
3.2	Funktionsprinzip	7
3.2.1	Attribut	7
3.2.2	Anfragen	8
3.3	Indizierung	9
3.4	Inzidenz-Matrix	9
3.5	Invertierte Liste	9
3.5.1	Verarbeitung einer Anfrage	10
<b>4</b>	<b>Das Vektorraum-Modell</b>	<b>11</b>
4.1	Funktionsprinzip	11
4.2	Term Frequency	11

---

4.3	Document Frequency .....	11
4.4	Inverted Document Frequency .....	11
4.5	$Tf \times idf$ Weighting .....	11
4.5.1	Formeln .....	11
4.6	Ähnlichkeitsfunktion .....	11
4.6.1	Euklidische Distanz .....	11
4.6.2	Cosine Similarity .....	11
4.6.3	Alternativen .....	11
<b>5</b>	<b>Bewertung eines Information Retrieval Systems</b> .....	<b>12</b>
5.1	Precision .....	12
5.2	Recall .....	12
<b>6</b>	<b>Implementierung</b> .....	<b>13</b>
<b>7</b>	<b>Die Benutzeroberfläche</b> .....	<b>14</b>
<b>8</b>	<b>Zusammenfassung und Ausblick</b> .....	<b>15</b>
	Literaturverzeichnis .....	16
	Glossar .....	17
	Erklärung der Kandidatin / des Kandidaten .....	18

---

## Abbildungsverzeichnis

---

## Tabellenverzeichnis

# Einleitung und Problemstellung

## 1.1 Einleitung

Nahezu jeder nutzt heutzutage E-Mail-Dienste und kennt die Problematik, dass der Posteingang sich unter der Flut täglich eintreffender Nachrichten stetig füllt, bis der Ordner unübersichtlich voll ist.

Dies wird zu einem ernsthaften Problem, sobald etwas bestimmtes darin wiedergefunden werden soll.

Die benötigte Mail war wichtig, weil sie eine bestimmte Kontaktadresse enthielt, aber wie war nochmal der Absender? Längst vergessen. Das genaue Datum? Leider ist nur noch der Monat bekannt. Wer die Suche manuell bewältigen muss, ist an dieser Stelle verloren.

Abhilfe schafft ein Information Retrieval System, mit dem man bestimmte Suchkriterien eingeben und beliebig miteinander kombinieren kann.

Das System liefert idealerweise eine Reihe passender Resultate und der Anwender braucht sich nicht selbst durch hunderte von Mails durchzuarbeiten.

In diesem Fall könnte die Person beispielsweise im Freitext nach dem Wort „*Kontaktadresse*“ suchen und weitere Kriterien wie „*Datum = Juni*“ hinzufügen.

Besonderheit ist das beliebige Verknüpfen: Der Anwender kann sich entscheiden, ob er nur Resultate akzeptiert, auf die Beides zutrifft, oder ob es bereits reicht, wenn eines der Kriterien erfüllt ist. Dies erlaubt eine sehr individuelle, auf die Informationsbedürfnisse des Nutzers zugeschnittene Suche.

Ein solches System ist nicht nur für das Alltagsbeispiel E-Mail-Ordner, d.h. Posteingang, Postausgang etc. wertvoll, sondern lässt sich auch auf jede andere Art von Dokumentenarchiv, dessen Dokumente sowohl Freitext als auch Metadaten beinhalten, anwenden.

## 1.2 Problemstellung

Ziel der Arbeit ist die Konzeption und Realisierung eines Systems zur Informationssuche (engl. Information Retrieval System) in einem Dokumentenarchiv, wobei die Dokumente von teilweise strukturierter Natur sind.

Dies bedeutet, dass sie sowohl gewöhnlichen Freitext als auch Metadaten enthalten. Der Nutzer soll spezifizieren können, in welchen Metadaten er suchen möchte,



zudem soll die Freitextsuche auswählbar sein. Alle Suchanfragen sollen hierbei beliebig mit den booleschen Operatoren *AND* (engl. und) sowie *OR* (engl. oder) verknüpfbar sein.

Hauptanwendungszweck des Systems sind E-Mail-Archive wie Posteingang und Postausgang, allerdings soll das System so flexibel sein, dass es auch auf andere teilweise strukturierte Dokumentenarchive anwendbar ist.

## 1.3 Teilprobleme

Aus der Aufgabenstellung ergeben sich die im folgenden beschriebenen Teilprobleme.

### 1.3.1 Dynamisches einlesen der Metadaten

Die genauen Metadaten sind, da das System flexibel sein soll, vor dem Ausführen des Systems noch nicht bekannt. Demnach muss das IR-System die Namen der Metadaten beim Starten des Programms dynamisch einlesen und diese dem Nutzer auf der grafischen Oberfläche anzeigen.

### 1.3.2 Unterscheidung Metadaten und Freitext

Damit das System die Metadaten dynamisch einlesen kann, müssen die folgenden Punkte erfüllt sein:

1. Das System muss zwischen Metadaten und Freitext unterscheiden können.
2. Metadaten setzen sich aus Name und Inhalt zusammen, weshalb beides erkannt und voneinander abgegrenzt werden muss. Im folgenden werden die Namen als „Keywords“ bezeichnet.
3. Der Inhalt kann unterschiedlichen Datentyps sein, z.B. String oder Liste, weshalb dieser bestimmt werden muss.

### 1.3.3 Metadatensuche

Es muss erkannt werden, welche Metadaten der Nutzer ausgewählt hat und in genau diesen muss, unter Berücksichtigung des jeweiligen Datentyps der Inhalte, gesucht werden.

Im Gegensatz zur Freitextsuche muss zu jedem Dokument vor der Suche zunächst geprüft werden, ob das entsprechende Schlüsselwort überhaupt darin auftritt.

### 1.3.4 Freitextsuche

Bei der Freitextsuche ist die Wortzahl weitaus größer als bei der Metadatensuche. Daraus resultieren zwei Probleme:

1. Wie kann effizient in großen Wortmengen gesucht werden?
2. Wie kann die Suche bei begrenztem Speicher bewältigt werden?

Zudem stellt sich die Frage nach einem geeigneten Verfahren, das bei längeren Anfragen auch teilweise passende Ergebnisse liefert und messen kann, wie gut die erzielten Treffer zur Nutzeranfrage passen.

### 1.3.5 Verknüpfung mit UND/ODER

Alle Anfragen sollen beliebig mit den logischen Operatoren *AND*, *OR* sowie *NOT* verknüpfbar sein. Dies beinhaltet die folgenden Problemstellungen:

1. Keywordsuche und Freitextsuche müssen miteinander verknüpft werden.
2. Sind mehrere Keywords ausgewählt, müssen die Teilergebnisse verknüpft werden.
3. Stellt der Nutzer mehrere Anfragen, müssen die Ergebnisse der einzelnen Anfragen verknüpft werden.

### 1.3.6 Benutzeroberfläche

Der Nutzer benötigt eine verständliche Benutzeroberfläche, die es ihm ermöglicht, seine Suchanfragen beliebig zusammenzustellen. Hierzu muss die Oberfläche folgende grundlegenden Funktionalitäten aufweisen:

1. Das Suchverzeichnis, d.h. das Dokumentenarchiv in welchem die Suche stattfindet, muss auswählbar sein.
2. Alle im Archiv auftretenden Keywords sowie die Freitextsuche müssen auswählbar sein.
3. Logische Operatoren (AND,OR,NOT) zur Verknüpfung müssen auswählbar sein.
4. Die Suchanfrage muss für den Nutzer verständlich angezeigt werden.

## Information Retrieval

Dieses Kapitel soll dem Leser einen Überblick über die Bedeutung des Begriffs „Information Retrieval“ vermitteln.

### 2.1 Bedeutung

Der englische Begriff „Information Retrieval“ lässt sich mit „Informationsrückgewinnung“ ins Deutsche übersetzen. [Aca] Hierbei wird explizit von *Rückgewinnung* gesprochen, da keine neuen Informationen erzeugt werden.

Bevor erklärt wird, wofür Informationsrückgewinnung tatsächlich steht, wird zunächst auf den Teilbegriff „Information“ eingegangen.

#### 2.1.1 Information

Die Bedeutung des Begriffs Information ist sehr weit gefasst, was eine einheitliche Definition unmöglich macht.

Er stammt aus dem Lateinischen (*informare* = Gestalt geben) und heißt im Übertragenen Sinne so viel wie jemanden durch Unterweisung bilden.

Dies betont den Aspekt, dass eine Information stets einen Empfänger besitzt, welcher „gebildet“ wird. Dies kann eine Person, aber auch ein geeignetes, nach außen wirksames System sein.

Daten als Träger von Informationen müssen vom Empfänger aufgenommen und korrekt interpretiert werden, damit aus den Daten tatsächlich Informationen entstehen.

Die Informationen müssen deshalb auf irgendeine Weise dargestellt werden, z.B. durch alphabetische Zeichen, außerdem muss es hierfür einen geeigneten Träger geben. Dies kann beispielsweise ein Textdokument sein.

Information lassen sich in die folgenden drei Bestandteile zerlegen:

- Syntaktischer Teil: Ist die Struktur der Information syntaktisch zulässig? Beispiel hierfür ist die Einhaltung von Rechtschreibung und Grammatik bei Texten.
- Semantischer Teil: Welchen inhaltliche Bedeutung besitzt die Information?
- Pragmatischer Teil: Welchem Zweck dient sie?

([PDVC06], S.314-315)

### 2.1.2 Information Retrieval

Nachdem bekannt ist, worum es sich bei Informationen handelt, wird im Folgenden beschrieben, worauf sich „Information Retrieval“ bezieht.

Auch hier ist es problematisch, eine einheitliche Definition zu finden. Eine mögliche Erklärung lautet so:

**Definition 2.1.** (*Information Retrieval*)

*Mit Information Retrieval, kurz IR, wird das Auffinden von in unstrukturierter Form vorliegender und ein Informationsbedürfnis befriedigender Materialien innerhalb großer Sammlungen bezeichnet.*

Mit unstrukturierten Materialien sind hierbei meist Dokumente in Textform gemeint. Üblicherweise liegen die Sammlungen auf dem Computer gespeichert vor ([CDM08], S.1).

### 2.1.3 Unterschied zur Datenbankensuche

Information Retrieval unterscheidet sich stark von der Suche in Datenbanken.

In einer Datenbank liegen die Daten strukturiert in Form von Werttupeln bekannten Datentyps vor, was Definition 2.1 widerspricht.

Im Gegensatz zum Information Retrieval kann bei der Datenbankensuche nicht mit vagen Anfragen umgegangen werden. In der Datenbank kann zwar nach (*Miete* < 300) gesucht werden, aber mit „*günstige Miete*“ wäre das System überfordert: Wie ist günstig zu interpretieren? ([Fer03], S.10)

Ein Information Retrieval System muss solche Anfragen mit unklarer Bedeutung verarbeiten können.

## 2.2 Beispiel Websuche

Zum besseren Verständnis soll an dieser Stelle ein Beispiel zur Veranschaulichung gegeben werden.

Nahezu jeder benutzt im Alltag Web-Suchmaschinen. Websuche stellt eine Form Information Retrieval dar, da hier Freitext beinhaltende Dokumente (z.B. im HTML- oder pdf-Format) aufgefunden werden sollen, um das Informationsbedürfnis des Internetnutzers zu befriedigen. ([Fer03], S.6)

Möchte dieser zum Beispiel seinen nächsten Urlaub planen, könnte seine Suchanfrage „*Hotel günstig Kreta*“ lauten.

Die gesuchten Informationen dienen also dem Zweck, den Urlaub zu planen.

Problematisch ist hierbei der semantische Teil der Informationen: Die inhaltliche Bedeutung der Resultate muss mit der ursprünglichen Intention des Nutzers

übereinstimmen.

Ein von der Suchmaschine geliefertes Resultat kann zwar zum syntaktischen Teil passen, da die korrekten Wörter darin auftauchen, allerdings in einem ganz anderen Kontext, sodass der Nutzer mit dem Dokument nichts anfangen kann.

## 2.3 Bezug zur Problemstellung

Die Aufgabe besteht darin, nach vom Nutzer auswählbaren, logisch verknüpften Kriterien innerhalb eines Dokumentenarchivs zu suchen (1.2).

Demnach ist die Definition 2.1 erfüllt, da hier Materialien innerhalb einer Sammlung, dem Dokumentenarchiv, aufgefunden werden sollen, um ein Informationsbedürfnis zu befriedigen.

Dieses Bedürfnis unterscheidet sich natürlich von Anfrage zu Anfrage, besteht aber allgemein gefasst darin, Dokumente wiederzufinden, z.B. eine bestimmte E-Mail.

### 2.3.1 Teilweise strukturierte Daten

Besonderheit der Problemstellung ist hierbei, dass die Dokumente teilweise strukturiert sind, d.h. es liegt zwar Freitext vor, aber zusätzlich sind Metadaten vorhanden.

Im Falle der Freitextsuche lässt sich aufgrund der unstrukturierten Textform eindeutig von Information Retrieval sprechen.

Anders sieht es bei den Metadaten aus, welche alle die folgende Syntax und damit Struktur besitzen:

(Name Inhalt)

Es liegt jedoch immer noch ein Information Retrieval Problem vor, da der Begriff auch die Suche in teilweise strukturierten (engl. semistructured) Dokumenten einschließt ([CDM08], S.1-2).

Wobei hierbei angemerkt sei, dass selbst die Metadaten nicht vollkommen strukturiert sind: Der Datentyp des Inhalts ist offen gelassen und es gibt keinerlei Vorgaben, welche Keywords in den Dokumenten auftreten müssen.

In den folgenden Kapiteln wird beschrieben, auf welche Weise ein Information Retrieval System konzipiert und realisiert werden kann, welches in der Lage ist, die Problemstellung 1.2 zu lösen.

Hierzu werden zunächst die hierfür benötigten Modelle „Boolean Retrieval“ sowie das „Vector Space Model“ erläutert.

## Boolesches Retrieval

Dieses Kapitel stellt das klassische Information-Retrieval-Verfahren „Boolesches Retrieval“ (engl. „Boolean Retrieval“) vor.

### 3.1 Eigenschaften des Verfahrens

Boolesches Retrieval überprüft Dokumente darauf, ob eine bestimmte Bedingung zutrifft.

Somit gibt es nur die Unterteilung in passende Dokumente und solche, welche die Bedingung nicht erfüllen. Eine weitere Bewertung der Ergebnisse findet nicht statt ([Fer03], S.33). Das fehlende Ranking der Ergebnisse ist ein häufiger Kritikpunkt des Verfahrens.

### 3.2 Funktionsprinzip

Boolesches Retrieval basiert auf Mengenoperationen. Deshalb werden Dokumente Mengen zugeordnet, die jeweils durch bestimmte Attribute charakterisiert sind.

Dokument bezeichnet die Einheit, auf welcher das Retrieval stattfindet. Ein Dokument kann deshalb eine kleine Textmemo, aber auch ein ganzes Buchkapitel sein ([CDM08], S.4).

#### 3.2.1 Attribut

Ein solches Attribut ist eine Abbildung, welche jedem Dokument einen Wert für dieses Attribut zuordnet. Die Abbildung erzeugt somit Attribut-Wert-Paare, was in Formel 3.1 gezeigt wird.

$$t : D \rightarrow T, t(d) = t_i \tag{3.1}$$

Hierbei bezeichnet  $t$  die Abbildung (d.h. das Attribut),  $D$  die Menge aller Dokumente und  $T$  den Wertebereich des Attributs  $t$ .

Der Attributwert  $t_i$  mit  $t_i \in T$  und  $i \in N$  wird durch die Abbildung  $t$  dem Dokument  $d \in D$  zugeordnet.

### 3.2.2 Anfragen

#### Elementare boolesche Anfrage

Ein Attribut-Wert-Paar wird auch als elementare boolesche Anfrage bezeichnet. Bei der elementaren booleschen Anfrage  $(t, t_1)$  werden zum Beispiel alle Dokumente gesucht, deren Attribut  $t$  den Wert  $t_1$  annimmt.

Mathematisch kann die Ergebnismenge  $D_{t,t_i}$  für eine Anfrage  $(t, t_i)$  wie in Formel 3.2 beschrieben werden.

$$D_{t,t_i} = \{d \in D \mid t(d) = t_i\} \quad (3.2)$$

#### Verknüpfung

Mehrere Attribut-Wert-Paare lassen sich mit booleschen Operatoren *AND*, *OR* und *NOT* verknüpfen.

$(t, t_1)$  *AND*  $(s, s_1)$  bedeutet, dass alle Dokumente gesucht sind, bei denen sowohl  $t(d) = t_1$  als auch  $s(d) = s_1$  gilt, d.h. hier muss der Durchschnitt dieser beiden Ergebnismengen gebildet werden, wie in Formel 3.3 gezeigt.

$$D_{t,t_1} \cap D_{s,s_1} \quad (3.3)$$

Wird hingegen der Operator *OR* verwendet, werden die Ergebnismengen vereinigt, wie in Formel 3.4 gezeigt.

$$D_{t,t_1} \cup D_{s,s_1} \quad (3.4)$$

Außerdem kann der unäre Operator *NOT* verwendet werden, der das Komplement der Ergebnismenge erzeugt. Demnach wird für die Anfrage *NOT*  $(t, t_1)$  erst die Menge aller Dokumente bestimmt, bei denen  $t(d) = t_1$  zutrifft, und diese anschließend von der Gesamtmenge aller Dokumente abgezogen. Dies wird in Formel 3.5 dargestellt.

$$D \setminus D_{t,t_1} \quad (3.5)$$

Da bei jeder Operation neue Ergebnismengen entstehen, lassen sich hierauf erneut die oben beschriebenen Operatoren anwenden. Auf diese Weise können Anfragen beliebig tief geschachtelt werden ([Fer03], S.34).

### 3.3 Indizierung

Um boolesche Anfragen zu verarbeiten zu können, müssen die Terme der Dokumente zunächst indiziert werden.

Mit „Term“ wird die Einheit, welche indiziert wird, bezeichnet. Dies kann ein Wort, aber auch eine andere Einheit wie die entsprechende Stammform sein.

Auch die Dokumente selbst müssen mit einem Index versehen werden.

Dieses Vorgehen ist unabdingbar, da ohne Indizierung nicht effizient auf die Dokument sowie deren Terme zugegriffen werden kann. Eine fehlende Indizierung würde bedeuten, dass für jede Anfrage alle Dokumente der Sammlung vollständig durchlaufen werden müssen, was unzumutbar langsam wäre ([CDM08], S.3).

### 3.4 Inzidenz-Matrix

Eine mögliche Implementierung des Boolean Retrieval stellt die Umsetzung mittels einer Term-Dokument-Inzidenz-Matrix dar. Dies bedeutet, dass die Zeilen der Matrix die Terme enthalten und die Spalten die Dokumente, was auch umgekehrt realisierbar ist. Tritt Term  $t$  in Dokument  $d$  auf, so lautet der Eintrag für  $(t, d)$  der Matrix 1. Alle Einträge für nicht vorkommende Terme sind hingegen mit einer 0 versehen ([CDM08], S.4).

Diese Herangehensweise ist belegt jedoch unnötig viel Speicherplatz, da sehr viele Einträge der Matrix eine 0 enthalten. Gerade bei sehr großen Sammlungen bzw. Dokumenten ist dies nicht realisierbar.

#### 3.4.1 Verarbeitung einer Anfrage

### 3.5 Invertierte Liste

In der Regel werden zur Implementierung invertierte Listen verwendet ([Fer03], S.36).

Der Name basiert auf den in den Listen gespeicherten „invertierten Indizes“. Die Indizes werden deshalb als invertiert bezeichnet, weil sie vom Term zurück auf die Position, in welcher der Term aufgetreten ist, schließen lassen.

Dies bedeutet, dass in einer geeigneten Speicherstruktur, zum Beispiel in einem Dictionary, zu jedem Term eine Liste gespeichert wird, die alle Dokumente enthält, in denen der Term auftritt.

Diese Dokumentlisten werden als invertierte Listen (engl. *inverted lists*) bezeichnet. Manche Implementierungen beinhalten auch zusätzliche Informationen wie die genaue Wortposition im Dokument.

Dieses Vorgehen setzt voraus, dass zuvor eine Indizierung (siehe 3.3) stattgefunden hat ([CDM08], S.5-6).

Hierbei sei noch angemerkt, dass die invertierten Listen auch in anderen Datenstrukturen als einer Liste gespeichert werden können.

Das Verfahren ermöglicht sehr schnelle Zugriffe, ist allerdings speicherintensiv ([Fer03], S.36). Im Vergleich zur Inzidenz-Matrix wird jedoch deutlich weniger Speicher benötigt, da die vielen leeren Einträge entfallen.



---

### 3.5.1 Verarbeitung einer Anfrage

## Das Vektorraum-Modell

### 4.1 Funktionsprinzip

### 4.2 Term Frequency

### 4.3 Document Frequency

### 4.4 Inverted Document Frequency

### 4.5 $Tf \times idf$ Weighting

#### 4.5.1 Formeln

### 4.6 Ähnlichkeitsfunktion

#### 4.6.1 Euklidische Distanz

#### 4.6.2 Cosine Similarity

#### 4.6.3 Alternativen

## Bewertung eines Information Retrieval Systems

### 5.1 Precisison

### 5.2 Recall

## Implementierung

## Die Benutzeroberfläche

## Zusammenfassung und Ausblick

In diesem Kapitel soll die Arbeit noch einmal kurz zusammengefasst werden. Insbesondere sollen die wesentlichen Ergebnisse Ihrer Arbeit herausgehoben werden. Erfahrungen, die z.B. Benutzer mit der Mensch-Maschine-Schnittstelle gemacht haben oder Ergebnisse von Leistungsmessungen sollen an dieser Stelle präsentiert werden. Sie können in diesem Kapitel auch die Ergebnisse oder das Arbeitsumfeld Ihrer Arbeit kritisch bewerten. Wünschenswerte Erweiterungen sollen als Hinweise auf weiterführende Arbeiten erwähnt werden.

---

## Literaturverzeichnis

- Aca.      *Academic - Academic dictionaries and encyclopedias, Universal Lexikon.*  
[http://universal\\_lexikon.deacademic.com/253489/Informationsr%C3%BCckgewinnung](http://universal_lexikon.deacademic.com/253489/Informationsr%C3%BCckgewinnung), letzter Zugriff am 05.06.2017.
- CDM08.   CHRISTOPHER D. MANNING, PRABHAKAR RAGHAVAN, HINRICH SCHÜTZE: *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- Fer03.    FERBER, REGINALD: *Information Retrieval, Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*. dpunkt.verlag, 2003.
- PDVC06.   PROF. DR. VOLKER CLAUS, PROF. DR. ANDREAS SCHWILL: *Duden - Informatik A-Z, Fachlexikon für Studium, Ausbildung und Beruf*. Dudenverlag, 2006.

# A

---

## Glossar

DisASter	DisASter (Distributed Algorithms Simulation Terrain), A platform for the Implementation of Distributed Algorithms
DSM	Distributed Shared Memory
AC	Linearisierbarkeit (atomic consistency)
SC	Sequentielle Konsistenz (sequential consistency)
WC	Schwache Konsistenz (weak consistency)
RC	Freigabekonsistenz (release consistency)



## B

---

### Erklärung der Kandidatin / des Kandidaten

☐ Die Arbeit habe ich selbstständig verfasst und keine anderen als die angegebenen Quellen- und Hilfsmittel verwendet.

☐ Die Arbeit wurde als Gruppenarbeit angefertigt. Meine eigene Leistung ist ...

Diesen Teil habe ich selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet.

Namen der Mitverfasser: ...

---

Datum

---

Unterschrift der Kandidatin / des Kandidaten