

Aufgabe 1

Gegeben sind folgende Dokumente

Doc1: Neues aus der Literaturwissenschaft: Durchbruch in der Shakespeareforschung.

Doc2: ein neues buch von shakespeare entdeckt.

Doc3: Freude für Shakespeareliebhaber.

Doc4: Waren das schon alle neuen Bücher Shakespeares?

- 1) Konstruieren Sie einen Index durch folgende Zwischenschritte:
Tokenisierung, Normalisierung(Lemmatisierung), Sortieren und Gruppieren.
- 2) Was liefern folgende booleschen Anfragen für die vier Dokumente in der vorherigen Aufgabe zurück:
 - neu AND Shakespeare
 - Buch AND neu
- 3) Wieweit hängen die Ergebnisse von der Normalisierung der Token (linguistische, Bearbeitung wie Lemmatisierung etc.) ab?
- 4) Wieweit hängen die Ergebnisse von der Tokenisierung ab? Was wäre das Ergebnis wenn ein Kompositasplitter eingebaut ist.

Aufgabe 2

(a) Was sind Vor- und Nachteile des Booleschen Retrievals?

(b) Angenommen, die folgenden Dokumente seien von einem Booleschen Retrievalsystem indexiert worden. Dabei fand die übliche Stoppworteliminierung, sowie Stemming statt.

1. Evaluating Strategic Support for Information Access in the Daffodil System.
 2. Daffodil: A User-Oriented Approach for Accessing Federated Digital Libraries.
 3. Daffodil: Distributed Agents for User-Friendly Access of Digital Libraries.
 4. Daffodil - Strategic Support for User-Oriented Access to Heterogeneous Digital Libraries.
 5. Active Support for Query Formulation in Virtual Digital Libraries: A case study with Daffodil.
 6. Daffodil: An integrated desktop for supporting high-level search activities in federated digital libraries.
 7. User-Oriented Query Modification in Metaclass Systems.
 8. Daffodil: Strategic Support Evaluated
- Formulieren Sie möglichst knappe Boolesche Anfragen (mit AND, OR und NOT), die genau die folgenden Dokumente finden:

(i) 2 und 4

(ii) 7

(iii) 8

Aufgabe 3: Computerlinguisten und Informatiker 9 ETCS

Finden Sie das Papier: „Monolingual Document Retrieval for European Languages“ von Vera Hollink et al im Web. Lesen Sie quer und schreiben Sie eine halbe Seite zu den Ergebnissen für Deutsch. Was wurde wie untersucht, was wäre Ihre Empfehlung wenn ein System implementiert werden soll.