

# Information Retrieval

## Übungen

Ingo Frommholz

Universität Duisburg-Essen

14.05.2008

UNIVERSITÄT  
DUISBURG  
ESSEN

# Klassifikation und Suche

Betrachte den Webkatalog unter <http://www.dmoz.org/>.

1. Untersuche die Klassifikation in Hinblick auf die in der Vorlesung genannten Merkmale und Kriterien. Um was für eine Art von Klassifikation handelt es sich? Versuche auch, Schwächen der konkreten Klassifikation herauszuarbeiten. (5 min.)

# Klassifikation und Suche

Betrachte den Webkatalog unter <http://www.dmoz.org/>.

1. Untersuche die Klassifikation in Hinblick auf die in der Vorlesung genannten Merkmale und Kriterien. Um was für eine Art von Klassifikation handelt es sich? Versuche auch, Schwächen der konkreten Klassifikation herauszuarbeiten. (5 min.)
2. Angenommen, Du suchst nach Dokumenten zu König Artus. Versuche zunächst diese über die Klassifikation zu finden. Wie gehst Du vor? Welche Probleme treten auf? (10 min)

# Klassifikation und Suche

Betrachte den Webkatalog unter <http://www.dmoz.org/>.

1. Untersuche die Klassifikation in Hinblick auf die in der Vorlesung genannten Merkmale und Kriterien. Um was für eine Art von Klassifikation handelt es sich? Versuche auch, Schwächen der konkreten Klassifikation herauszuarbeiten. (5 min.)
2. Angenommen, Du suchst nach Dokumenten zu König Artus. Versuche zunächst diese über die Klassifikation zu finden. Wie gehst Du vor? Welche Probleme treten auf? (10 min)
3. Benutze nun die Suche. Welche Dokumente hast Du bei der Suche zusätzlich gefunden, bzw. welche Dokumente aus der Klassifikation fehlen im Ergebnis der Suche? Welches Herangehen war insgesamt erfolgreicher?

- Welche grundsätzlichen Probleme ergeben sich bei der Freitextsuche?

- Welche grundsätzlichen Probleme ergeben sich bei der Freitextsuche?
- Was sind ihre Auswirkungen auf Recall und Precision?

- Welche grundsätzlichen Probleme ergeben sich bei der Freitextsuche?
- Was sind ihre Auswirkungen auf Recall und Precision?
- Wie kann Abhilfe geschaffen werden?

- Welche grundsätzlichen Probleme ergeben sich bei der Freitextsuche?
- Was sind ihre Auswirkungen auf Recall und Precision?
- Wie kann Abhilfe geschaffen werden?

# Informatischer Ansatz

Gegeben sei folgender Text (aus:

[http://en.wikipedia.org/wiki/Monty\\_Python\\_and\\_the\\_Holy\\_Grail](http://en.wikipedia.org/wiki/Monty_Python_and_the_Holy_Grail)):

*King Arthur is recruiting his Knights of the Round Table throughout England. He is frustrated at every turn by anarcho-syndicalist peasants, a Black Knight that refuses to give up despite losing both his arms and legs, and guards that are more concerned with the flight patterns of swallows than their lord and master.*

Die folgende Anfrage sei gegeben: `Arthur recrute knight round table`. Wie muss die Anfrage bearbeitet werden, damit sowohl obiger Text als auch ein Text mit dem Satz "King Arthur recrutes another knight of the round table" gefunden wird? Dabei wird die Suche hier als Zeichenkettensuche verstanden.

# Computerlinguistischer Ansatz: Stoppworteliminierung und Grundformreduktion

Gegeben sei folgender Text (aus:  
[http://en.wikipedia.org/wiki/Monty\\_Python\\_and\\_the\\_Holy\\_Grail](http://en.wikipedia.org/wiki/Monty_Python_and_the_Holy_Grail)):

*King Arthur is recruiting his Knights of the Round Table throughout England. He is frustrated at every turn by anarcho-syndicalist peasants, a Black Knight that refuses to give up despite losing both his arms and legs, and guards that are more concerned with the flight patterns of swallows than their lord and master.*

Führe eine Stoppworteliminierung und anschließend eine Grundformreduktion nach Kuhlen durch. Was bleibt übrig vom Text?

Benutze die graphische Notation aus der Vorlesung, um die folgenden Aussagen in RDF-Schema zu modellieren:

*Kraftfahrzeuge sind auf Personen zugelassen. PKW, Bus und Mini-Van sind Arten von Kraftfahrzeugen. Ein Kraftfahrzeug hat eine Marke, eine Farbe und ein Kennzeichen. Eine Person hat einen Namen und eine Ausweisnummer.*

Modelliere zum RDF-Schema die folgenden Aussagen:

*Der rote GMC Savana mit dem Kennzeichen 'E-DU 1111' ist ein Mini-Van. Der Wagen ist auf Hans Müller zugelassen.*