

# Modul IRG : Semesterkurzbeitrag

## Ausgangslage

Im Rahmen Ihres Semesterkurzbeitrags werden Sie 50 Anfragen für eine Kollektion von über 20'000 Dokumenten bearbeiten. Die Anfragen stellen Informationsbedürfnisse dar, welche einerseits in Stichworten, aber auch als vollständige Sätze formuliert sind. Sie werden Information Retrieval Systeme auf ihre Fähigkeit hin untersuchen, diese Informationsbedürfnisse innerhalb der Dokumentenkollektion zu befriedigen.

Die Dokumentenkollektion setzt sich aus drei häufig in der Forschung verwendeten Kollektionen zusammen. Es handelt sich dabei um die Titel und Zusammenfassungen von wissenschaftlichen Informatikpublikationen.

## Aufgaben

Ihre Aufgabe ist es, für die Anfragen möglichst gute oder interessante Suchresultate in der Kollektion zu finden. Sie erstellen mindestens eine Datei, welche für jede Anfrage eine Rangliste von Suchresultaten enthält. Dazu schreiben Sie einen Bericht, welcher Ihren Lösungsweg aufzeigt.

## Rangliste

Sie geben eine Ranglistendatei im trec\_eval Format ab, welche je 1000 Suchresultate für alle 50 Anfragen enthält. Wählen Sie die in Ihren Augen beste oder interessanteste Rangliste aus.

Das trec\_eval Dateiformat enthält eine Liste von Zeilen mit folgendem Aufbau:

Spalte	Query ID	Iteration	Dok.Nummer	Rang	Score	System
Inhalt	ID der Anfrage	Konstant auf Q0 (Qnull)	ID eines gefundenen Dokuments	Rang des Dokuments	RSV	Name des Systems, verwenden Sie Ihren Gruppen-namen
Datentyp	Integer	String	String	Integer	Float	String

Die Felder werden jeweils durch ein einzelnes Leerzeichen getrennt. Beispielzeilen:

```
30 Q0 ZF08-175-870 0 4.238 inrg1
30 Q0 KS09-937-830 1 4.097 inrg1
30 Q0 OW01-739-219 2 3.951 inrg1
```

Nach der Abgabe aller Beiträge werden die Ranglisten verglichen und öffentlich publiziert.

## Bericht

Verfassen Sie einen Bericht, welcher lesbar auf maximal zwei A4 Seiten Platz hat. Darin erklären Sie ihre abgegebene Rangliste. Führen Sie aus, welche Systeme (Beispiele: Eigenbau, Laborumgebung, Google Desktop Search, Microsoft Search Server, etc.) Sie eingesetzt haben und was die Motivation war. Beschreiben Sie den Aufbau des Index, welchen Sie erzeugt haben. Zeigen Sie, wie die Anfragen an die Systeme gestellt wurden (manuell, automatisch, iterativ, etc.) und wie Sie die Resultate interpretieren. Wie die Ranglisten werden auch die Berichte öffentlich aufgeschaltet.

Die Qualität dieses Berichts resp. die fachliche Individualität der Arbeit ist die Grundlage für Ihre Bewertung. Dabei sind Sie nicht darauf beschränkt, nur möglichst „gute“ Suchresultate zu erhalten. Auch experimentelle Ansätze wie zum Beispiel Optimierung auf Antwortgeschwindigkeit sind

willkommen, sofern die Resultate interessant sind und brauchbar bleiben. Ihre Suchresultate validieren die Aussagen Ihres Berichts.

## Rahmenbedingungen

- Arbeit in **Zweiergruppen**: Organisieren Sie sich frühzeitig, damit Sie bald loslegen können.
- Ausgabe am Dienstag, 19. Februar 2013
- Ein **kurzes "Konzept"** (ca. 10 Zeilen) per **Email an** bram@zhaw.ch **bis 15. März 2013**
- In den **Praktika vom 2. April** findet eine kurze **Fragestunde** zum Verlauf der Semesterkurzbeiträge statt (max. 1 Lektion).
- In den Praktika wird immer wieder auf "Überlappungen" mit der Aufgabenstellung der Semesterkurzbeiträge hingewiesen.
- **Abgabe** bis am **Freitag, 8. Mai 2013 um Mitternacht**
  - Überprüfen Sie das **Format der Rangliste** sorgfältig
  - Grösse des **Berichts möglichst 2 A4 Seiten**
- Die **Bewertung** erfolgt nach diesem Massstab (Punkte werden an Semesterendprüfung angerechnet):
  - 0 Punkte für eine minimalistische Arbeit
  - 5 Punkte für eine minimalistische Arbeit, z.B. ein vorgefertigtes System zum Laufen gebracht
  - 10 Punkte für eine vernünftige Auseinandersetzung mit dem Thema
  - 20 Punkte für eine sehr aufwändige und interessante Arbeit

## Lernziele

Sie sollen sich neben den regulären Praktikumsaufgaben auf eigene und kreative Weise mit dem Thema Information Retrieval auseinandersetzen. Mit Ihrem Wissensaufbau im Modul lernen Sie jede Woche wieder neue Ansätze kennen, welche Sie sukzessive in den Semesterkurzbeitrag einfließen lassen können. Durch die systematische Betrachtung verschiedener Systeme sollen Sie unter anderem die Fähigkeit erlangen, in einer Firma Information Retrieval Systeme für den Gebrauch evaluieren zu können.

## Ressourcen

Die Kollektion und Anfragen beziehen Sie im OLAT System unter folgendem Link:  
<https://olat.zhaw.ch/olat/auth/1%3A1%3A0%3A0%3A0/>