



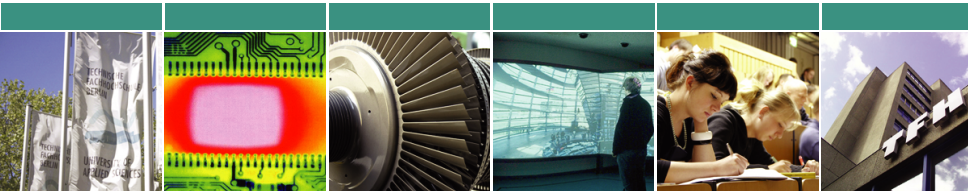
Datenbanksysteme – Übung

Technische Informatik (Bachelor)

1

Aljoscha Marcel Everding

Beuth Hochschule Berlin





① Organisatorisches

- ① Vorstellung
- ② Gruppenbildung
- ③ Regularien

② Inhalt

- ① Vorstellung der Aufgaben
- ② AOL Search Leak
- ③ Vorstellung von zwei Beispiellösungen

③ Zusammenfassung

Teil I

Organisatorisches

- Aljoscha Marcel Everding
- Gastdozent
- Freier Softwareentwickler und IT-Trainer

Kontakt

- Moodle
- E-Mail: aljoschamarcel.everding@beuth-hochschule.de
 - GnuPG-Key ID AC3FDFD4
 - Fingerprint 3811 DC50 C4D6 82A2 E78B 1119 C276 3073 AC3F DFD4
- Persönlich: an der Beuth-Hochschule
Di 08:15 – 19:15
Mi 08:15 – 15:45

- Aljoscha Marcel Everding
- Gastdozent
- Freier Softwareentwickler und IT-Trainer

Kontakt

- Moodle
- E-Mail: aljoschamarcel.everding@beuth-hochschule.de
 - GnuPG-Key ID AC3FDFD4
 - Fingerprint 3811 DC50 C4D6 82A2 E78B 1119 C276 3073 AC3F DFD4
- Persönlich: an der Beuth-Hochschule
Di 08:15 – 19:15
Mi 08:15 – 15:45



Gruppeneinteilung

Die Übungsaufgaben müssen von Ihnen als Gruppe (max. 4 Personen) bearbeitet und abgegeben werden. Die Einteilung erfolgt durchs Zufallsprinzip.

Datenerfassung

Erfassung der E-Mail-Adressen und Gruppen



- Übungen wöchentlich: 2 Gruppen gleichen Inhalts, jeweils Dienstags
- 4 bewertete Übungen (gesamt 30 Punkte)
 - Gruppenarbeit (max. 4 Personen) bis Semesterende!
- Klausur mit insgesamt 70 Punkten
- Die Gesamtnote resultiert aus den **insgesamt** erreichten Punkten (aus Klausur **und** den Übungen).
- **Übung erfolgreich:** bei **min. 15 Punkten** erreicht.

- Ankündigungen, Unterlagen, Übungsaufgaben
 - Anmelden unter: <http://lms.beuth-hochschule.de/>
 - Fachbereich VI → Medieninformatik Bachelor → Datenbanksysteme
 - Bitte Eintrag mit Nachname, Vorname, da Punkte- und Notenlisten daraus erzeugt werden
 - Aktivierung im gesamten Semester
 - Nach Nachklausur evtl. Deaktivierung

Teil II

Inhalt



- Sie wissen, was Sie in den Aufgaben erwartet.
- Sie haben mindestens Aufgabe 1 verstanden.
- Sie haben festgestellt, dass drei weitere Aufgaben auf Aufgabe 1 aufsetzen.

Grundlegende Datenbasis für die Aufgaben: AOL Search Leak¹

Overview [\[edit\]](#)

On August 4, 2006, AOL Research, headed by Dr. Abdur Chowdhury, released a compressed text file on one of its websites containing twenty million search [keywords](#) for over 650,000 users over a 3-month period, intended for research purposes. AOL deleted the search data on their site by the 7th, but not before it had been mirrored and distributed on the Internet.

AOL themselves did not identify users in the report; however, [personally identifiable information](#) was present in many of the queries, and as the queries were attributed by AOL to particular user accounts, identified numerically, an individual could be identified and matched to their account and search history by such information.^[1] *The New York Times* was able to locate an individual from the released and anonymized search records by cross referencing them with phonebook listings.^[2] Consequently, the ethical implications of using this data for research are under debate.^{[3][4]}

AOL acknowledged it was a mistake and removed the data. However, the removal was too late; the data was redistributed by others, and can still be downloaded from mirror sites.^{[5][6]}

In January 2007, Business 2.0 Magazine on CNNMoney ranked the release of the search data #57 in a segment called "101 Dumbest Moments in Business."^[7]

Screenshot http://en.wikipedia.org/wiki/AOL_search_data_leak
(abgerufen 06.04.2015)

Weiterführende Informationen auch unter <https://search-id.com/>

¹http://en.wikipedia.org/wiki/AOL_search_data_leak

1 Aufgabe 1: „Anfragen und Modellierung“

Denken Sie mal darüber nach, welche Anfragen Sie an die AOL Daten stellen möchten. Bitte modellieren Sie ein logisches und physisches Schema zur Beantwortung dieser Anfragen.

2 Aufgabe 2 „SQL- und Abfrageausführung“

Bitte formulieren Sie für Ihre Analyseideen aus 1.) die SQL-Anfragen. Sie verstehen auch Möglichkeiten der Abfrageausführung bzw. Optimierung.

3 Aufgabe 3 „Datenintegration“

Zur Ausführung fehlen Ihnen noch externe Daten, z.B. aus dem Internet Archive, DMOZ oder Freebase.org. Bitte ergänzen Sie Ihr Schema und die Datenbasis.

4 Aufgabe 4 „Analyse, Erkenntnisgewinn und Wert“

Stellen Sie in 10 Minuten die wichtigsten Erkenntnisse aus den Daten vor. Bewerten Sie den Erkenntnisgewinn, z.B. gegenüber Ihren Kommilitonen oder der Literatur! Welche Erkenntnisse hätten einen kommerziellen Wert?

1 Aufgabe 1: „Anfragen und Modellierung“

Denken Sie mal darüber nach, welche Anfragen Sie an die AOL Daten stellen möchten. Bitte modellieren Sie ein logisches und physisches Schema zur Beantwortung dieser Anfragen.

2 Aufgabe 2 „SQL- und Abfrageausführung“

Bitte formulieren Sie für Ihre Analyseideen aus 1.) die SQL-Anfragen. Sie verstehen auch Möglichkeiten der Abfrageausführung bzw. Optimierung.

3 Aufgabe 3 „Datenintegration“

Zur Ausführung fehlen Ihnen noch externe Daten, z.B. aus dem Internet Archive, DMOZ oder Freebase.org. Bitte ergänzen Sie Ihr Schema und die Datenbasis.

4 Aufgabe 4 „Analyse, Erkenntnisgewinn und Wert“

Stellen Sie in 10 Minuten die wichtigsten Erkenntnisse aus den Daten vor. Bewerten Sie den Erkenntnisgewinn, z.B. gegenüber Ihren Kommilitonen oder der Literatur! Welche Erkenntnisse hätten einen kommerziellen Wert?

**1 Aufgabe 1: „Anfragen und Modellierung“**

Denken Sie mal darüber nach, welche Anfragen Sie an die AOL Daten stellen möchten. Bitte modellieren Sie ein logisches und physisches Schema zur Beantwortung dieser Anfragen.

2 Aufgabe 2 „SQL- und Abfrageausführung“

Bitte formulieren Sie für Ihre Analyseideen aus 1.) die SQL-Anfragen. Sie verstehen auch Möglichkeiten der Abfrageausführung bzw. Optimierung.

3 Aufgabe 3 „Datenintegration“

Zur Ausführung fehlen Ihnen noch externe Daten, z.B. aus dem Internet Archive, DMOZ oder Freebase.org. Bitte ergänzen Sie Ihr Schema und die Datenbasis.

4 Aufgabe 4 „Analyse, Erkenntnisgewinn und Wert“

Stellen Sie in 10 Minuten die wichtigsten Erkenntnisse aus den Daten vor. Bewerten Sie den Erkenntnisgewinn, z.B. gegenüber Ihren Kommilitonen oder der Literatur! Welche Erkenntnisse hätten einen kommerziellen Wert?

1 Aufgabe 1: „Anfragen und Modellierung“

Denken Sie mal darüber nach, welche Anfragen Sie an die AOL Daten stellen möchten. Bitte modellieren Sie ein logisches und physisches Schema zur Beantwortung dieser Anfragen.

2 Aufgabe 2 „SQL- und Abfrageausführung“

Bitte formulieren Sie für Ihre Analyseideen aus 1.) die SQL-Anfragen. Sie verstehen auch Möglichkeiten der Abfrageausführung bzw. Optimierung.

3 Aufgabe 3 „Datenintegration“

Zur Ausführung fehlen Ihnen noch externe Daten, z.B. aus dem Internet Archive, DMOZ oder Freebase.org. Bitte ergänzen Sie Ihr Schema und die Datenbasis.

4 Aufgabe 4 „Analyse, Erkenntnisgewinn und Wert“

Stellen Sie in 10 Minuten die wichtigsten Erkenntnisse aus den Daten vor. Bewerten Sie den Erkenntnisgewinn, z.B. gegenüber Ihren Kommilitonen oder der Literatur! Welche Erkenntnisse hätten einen kommerziellen Wert?



Vorstellung exemplarischer Lösungen aus früheren Semestern:

- Oscar-Nominierungen (Viola Jertschat_285307_-
assignsubmission_file_A4_faruhn_jertschat_ngo_khong.pdf)
- Tornado
Anna-Sophia Wirth_282003_assignsubmission_file_-
Analyse-Verwertung_BenteGade_SophieWirth.pdf

Vorstellung exemplarischer Lösungen aus früheren Semestern:

- Oscar-Nominierungen (Viola Jertschat_285307_-
assignsubmission_file_A4_faruhn_jertschat_ngo_khong.pdf)
- Tornado
Anna-Sophia Wirth_282003_assignsubmission_file_-
Analyse-Verwertung_BenteGade_SophieWirth.pdf

Teil III

Zusammenfassung (TL;DR)



TL;DR

- Es müssen vier **aufeinander aufbauende** Übungsaufgaben in Gruppenarbeit erstellt und abgegeben werden.
- Grundlegende Datenbasis für die Aufgaben ist der *AOL Search Leak*.
- Der Umfang und die Art der Lösungen wurde an zwei Beispielen vorgestellt.



(by Duncan Hull, <https://www.flickr.com/photos/dullhunk/202872717/>)



(by Maik Meid, <https://www.flickr.com/photos/frnetz/7077212593/>)



This work is licensed under the *Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License*.