

Klausur

Katja Markert

January 28, 2020

- Dieser FAQ kann bis zum Tag vor der Klausur neue Informationen bekommen und (leicht) geändert werden.
- Nicht alle Eventualitäten sind angesprochen
- Viele Listen sind nicht vollständig

- **12.2.2019, 10.00 s.t. bis 12.45 (bitte um 10.00 da sein, damit wir um 10.15 anfangen können zu schreiben)**
- **HS2, INF306** (bitte Raum und Zeit kurz vor Klausur noch einmal Checken)
- Nur Studis, die auch committed sind: Commitmentfrist kommt noch.
- Nur Studis, die dieses Jahr oder in Vorgängerjahren die Zulassung (mind. 60% der Übungsaufgaben) erlangt haben
- Studis, die sich trotz unzureichender Zulassung committed haben, werden evtl. vorher rausgefiltert (obwohl nicht meine Aufgabe oder die des Sekretariats). Falls nicht, oder wenn jemand die Klausur einfach trotzdem mitschreibt, gilt er als durchgefallen!

- Sitzplätze horizontal mit mind. einem Platz frei zu benachbarten Studenten einnehmen.
- Laptops und Smartphones am besten zu Hause lassen. Sonst vorne an der Tafel ablegen. Alle Taschen bleiben geschlossen.

- Sitzplätze horizontal mit mind. einem Platz frei zu benachbarten Studenten einnehmen.
- Laptops und Smartphones am besten zu Hause lassen. Sonst vorne an der Tafel ablegen. Alle Taschen bleiben geschlossen.
- Auf die Klausur bitte **deutlich** auf Deckblatt Namen sowie Matrikelnummer schreiben. Studierendenausweis bitte neben sich auf Tisch legen.
- Beantwortung der Klausur auf leeren Zusatzbögen, die gestellt werden. Diese bitte auch deutlich mit Namen und Matrikelnummer versehen.

- Sitzplätze horizontal mit mind. einem Platz frei zu benachbarten Studenten einnehmen.
- Laptops und Smartphones am besten zu Hause lassen. Sonst vorne an der Tafel ablegen. Alle Taschen bleiben geschlossen.
- Auf die Klausur bitte **deutlich** auf Deckblatt Namen sowie Matrikelnummer schreiben. Studierendenausweis bitte neben sich auf Tisch legen.
- Beantwortung der Klausur auf leeren Zusatzbögen, die gestellt werden. Diese bitte auch deutlich mit Namen und Matrikelnummer versehen.
- **Keine** Zusammenfassung ("Spickzettel") oder weitere Unterlagen erlaubt.
- Erlaubt: Stifte (bitte nur blau oder schwarz, keine Bleistifte). Taschenrechner. Keine Lexika.

- Mit Beginn der Klausur bestätigen Sie Ihre geistige und physische Fitness.
- Bei Krankheit muss man sich vorher überlegen, ob man mitschreibt. Wenn nicht, ärztliches Attest vom vorherigem oder gleichen Tag zwingend, sonst gilt man als durchgefallen
- Bei plötzlicher Krankheit während Klausur: bitte sofort melden.
- Man darf früher abgeben, wenn man früher fertig ist und gehen. Allerdings nicht in den letzten 20 Minuten der Klausur.

- 5 Fragen, von denen 4 beantwortet werden müssen. D.H. eine Frage kann nach persönlichem Wunsch weggelassen werden.
Das Deckblatt ist zum Lesen da!!!
- Bitte schreiben Sie in das vorgesehene Feld auf dem Deckblatt deutlich, welche Fragen beantwortet wurden. Beispiel (bei Auswahl aus 5 Fragen).

Fragennummer	Punkte
1	von Dozent auszufüllen
3	von Dozent auszufüllen
4	von Dozent auszufüllen
5	von Dozent auszufüllen

- Fehlt diese Info auf dem Deckblatt, werden die ersten 4 Fragen, die mindestens ein (nicht-durchgestrichenes) geschriebenes Wort oder Formel/Zahl (token) enthalten korrigiert.
- Ich werde **nicht** alle Fragen korrigieren und dann das für Sie beste herausuchen. Sie müssen klar angeben, welche Fragen Sie gewählt haben.

- **50% zum Bestehen nötig**
- Folgefehler werden so wenig wie möglich bestraft, falls klare Folgefehler
- Von mehreren gegebenen (nicht durchgestrichenen) Antwortmöglichkeiten wird immer die **erste** bewertet. Wieder: ich suche nicht die beste Zusammensetzung für Sie heraus.
- Beispiel: Sollte eine Frage lauten “Beschreiben Sie zwei Nachteile dieses Verfahrens” und Sie nennen mir drei, dann werde ich die ersten beiden bewerten. Das Gießkannenverfahren ist also unerwünscht.

- Endergebnisse sollten so weit wie möglich vereinfacht werden, müssen aber nicht in den Taschenrechner eingegeben werden. Beispiel eines völlig legitimen Endergebnisses:

$$\cos(x, y) = \dots \frac{\sqrt{15}}{\sqrt{34}}$$

- Für Zwischenschritte gilt dies oft nicht. Beispiel: Wenn Sie bei einem Prob. CKY nicht die Zwischenwahrscheinlichkeiten ausrechnen, dann wird es problematisch...

- Formeln: Die wichtigsten und meistgebrauchten Formeln müssen Sie einfach kennen. Beispiele: precision, recall, F-measure, IR evaluation measures, Naive Bayes, cosine, PMI, N-gram, Bayes rule, K-means ...
- Formeln: Seltener gebrauchte oder nicht so prominente Formeln (i) kommen entweder nicht vor oder (ii) werden angegeben oder (iii) werden soweit erklärt, dass man Sie sich herleiten kann (Gedächtnisstütze)

Berechnen Sie für die unten angegebenen Vektoren ein 2-means clustering bis zur Konvergenz. Wie gut ist das Clustering, wobei Sie Clusterreinheit der einzelnen Cluster berechnen sollten?

- 1 Erste Alternative: Die Formel für Clusterreinheit ist:
- 2 Zweite Alternative: Zur Erinnerung: Clusterreinheit misst den maximalen Überlapp von einem Systemcluster mit einem Goldcluster

- Cluster purity
- MUC und B-Cubed für Koreferenz (2019-20 nicht besprochen)
- Smoothing außer Laplacesmoothing: also Interpolationssmoothing, Kneser-Ney etc
- Mutual Information (aber nicht für pointwise mutual information)
- PageRankformel zur Matrixumwandlung
- Manchmal könnte eventuell auch an Definitionen anstatt an Formeln erinnert werden.

- Alles von Anfang der Vorlesungszeit bis zum Ende des Themas *Information Retrieval*.
- Kein NLTK/Programmieren

Fragearten: “Rechenaufgaben/Simulationsaufgaben”

Diese Liste ist nicht vollständig!!!!

- Ähnlich zu Übungsaufgaben: berechnen Sie Clustering, Pagerank, Evaluationsmaße, n-gram Modell, minimum edit distance, Viterbi, PCFG, CKY, CNF, (multinomial/binomial) NB, FSA schreiben/lesen, RE schreiben/lesen ...
- Falls dies in manchen Fällen zu umfangreich: Gebe (partielle) Berechnung vor.
 - Vervollständige diese
 - Finde Fehler in vorgegebener Rechnung
 - Illustriere etwas an vorgegebener Rechnung
 - Erkläre Grundprinzipien an vorgegebener Rechnung
 - Was passiert, wenn ...

Diese Liste ist nicht vollständig!

- Gebe linguistische Daten vor. “Formalisiere” diese. Beispiele:
 - Tagge oder Parse
 - Erkläre linguistische Begriffe an diesen Daten (Synonymie, Ambiguität ...)
 - Leite Grammatik oder prob. Grammatik ab
 - Erkläre Probleme für ein bestimmtes Verfahren anhand der Daten
 - Erkläre Features anhand der Daten
 - ...
- Definitionsfragen (selten, aber möglich)
- Auch hier wieder gibt es Lösungsmöglichkeiten, falls zu umfangreich oder zu erinnerungsintensiv (z.B. alle Hockett Features für Sprachen): mehrere vorgeben und Studi darf drei aussuchen, die er erklärt/anwendet; vorgeben einer Analyse und Fehler finden etc.

Diese Liste ist nicht vollständig!

- Definitionsfragen (selten)
- Algorithmusprobleme erklären oder Algorithmen vergleichen (winziges Essay)
- Kleine Beweise oder Herleitungen
- Was-passiert-wenn Fragen
- Transferfragen zu neuen Problemen (siehe nächste Folie)
- Fragen zu kognitivem Hintergrund

Siehe hochgeladenes Klausur Beiblatt
Besprechung jetzt!

- Ich werde Studentenantworten auf Englisch oder Deutsch korrigieren. Auch gemischt, aber bitte nicht innerhalb einer Frage!
- Linguistische Beispiele/Daten können entweder auf Deutsch oder Englisch vorkommen (oder in Kunstsprachen). Manchmal führt Englisch zu einer einfacheren Aufgabe (Grammatik...). Manche Algorithmen wurden nur für Englisch entwickelt. Wenn beides nicht der Fall, wird versucht, deutsche Aufgaben zu stellen (oder in einer Kunstsprache).
- Falls englische Beispiele: Schwieriges Vokabular wird vermieden. Sollte die Bedeutung einer Vokabel unbekannt sein, kann man in der Klausur nachfragen.

- Übungsaufgaben (schauen Sie auch mal in das Übungsaufgabenset in Jurafsky und Martin)
- Zu jedem Task: Definition, Hauptprobleme, Datensätze, Evaluation, Hauptansätze
- Zu jedem Algorithmus/Ansatz/Merkmalssatz: Grundidee, Vorteile, Nachteile, Evaluation
- Slides und angegebene Literatur (mit Fokus auf Lehrbüchern, falls zu dem Thema Kapitel existent)
- Legen Sie Fokus auf die zentralen, lange besprochenen Algorithmen und Themen

Don't practice until you get it right, practice until you can't get it wrong.