

COMP6714 Review

Assignment Project Exam Help

Wei Wang

<https://powcoder.com>

School of Computer Science and Engineering
Universities of New South Wales

Add WeChat powcoder

November 11, 2020

Course Logistics

- ▶ **THE** formula:

$$\text{mark} = \begin{cases} 0.25 \cdot (\text{ass1} + \text{proj1}) + 0.50 \cdot \text{exam} & , \text{if } \text{exam} \geq 40 \\ 39FL & , \text{otherwise.} \end{cases}$$

- ▶ Exam date: Exact time to be announced, 2 Dec (Wed) afternoon.

- ▶ Pre-exam consultations:

- ▶ TBA

- ▶ TBA

- ▶ Sample exam papers to be released soon.

- ▶ Course survey or private messages to me on the forum.

(1) The final exam mark is important and you must achieve at least 40! (2) Supplementary exam is **only** for those who cannot attend final exam. (3) Apply for UNSW Special Consideration (SC) with sufficient evidence and the SC team will make the final decision.

About the Final Exam

Assignment Project Exam Help

- ▶ **Time:** 10 minutes reading time + 2 hr **open book** exam + 15 minutes **scanning+uploading+submission** time.
 - ▶ **Very important** for you to know how to scan, upload, and submit.

Practice before hand !! We will launch a practice session before hand.

<https://powcoder.com>

- ▶ Designed to test your *understanding* and familiarity of the core contents of the course.
- ▶ 100 (8 questions)
 - ▶ Similar to those in the assignment.

Add WeChat powcoder

Special Note on the Final Exam

Assignment Project Exam Help

- ▶ We trust every student will uphold the academic integrity.
- ▶ Severe consequences for any misconduct in the final exam.

<https://powcoder.com>

Add WeChat powcoder

About the Final Exam ...

Assignment Project Exam Help

- ▶ Read the instructions carefully.
- ▶ You can answer the questions in *any* order.
- ▶ Some of the “Advanced” Methods/algorithms/systems are not required, unless explicitly mentioned here.

Tip: *Write down intermediate steps, so that we can give you partial marks even if the final answer is wrong.*

Disclaimer: *We will go through the main contents of each lecture. However, note that it is by no means exhaustive.*

Assignment Project Exam Help

- ▶ incidence vector
- ▶ semantics of the query model (AND/OR/NOT, and other operators, e.g., /k, /S)
- ▶ inverted index, positional inverted index
- ▶ query processing methods for basic and advanced boolean queries (including phrase query, queries with /S operator, etc.)
- ▶ query optimization methods (list merge order, skip pointers)
- ▶ **Not required:** next-word index

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

- ▶ typical preprocessing steps: tokenization, stopword removal, stemming, lemmatization

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

- ▶ Why we need dedicated algorithms to build the index?
- ▶ ESBt: Blocked sort-based indexing
- ▶ SPIM1: Single-pass in-memory indexing
- ▶ Dynamic indexing: Immediate merge, no merge, logarithmic merge

<https://powcoder.com>

Add WeChat powcoder

- ▶ What is/why ranked retrieval?
- ▶ raw and normalized tf, idf
- ▶ cosine similarity
- ▶ tf-idf variants (using SMART notation): e.g., Inc.ltc
- ▶ basic query processing method: document-at-a-time vs term-at-a-time
- ▶ exact & approximate query optimization methods (heap-based top-k algorithm, MaxScore and WAND algorithms, etc.)
- ▶ **Not required:** Query processing methods based on advanced or tiered inverted indexes (e.g., high/low lists, impact-oriented lists, etc.)

Assignment Project Exam Help

- ▶ Existing method to prepare for the benchmark dataset, queries, and ground truth
- ▶ For unranked results: Precision, recall, F-measure
- ▶ For ranked results: precision-recall graph, 11-point interpolated precision, MAP, etc.
- ▶ **Not required:** NDCG, Kappa (κ) measure for inter-judge (dis)agreement

<https://powcoder.com>
Add WeChat powcoder

Assignment Project Exam Help

- ▶ Probability ranking principle (intuitively, how to rank documents and when to stop)
- ▶ derivation of the ranking formula of the probabilistic model
- ▶ the BM25 method
- ▶ Query-likelihood *unigram* language model with *Jelinek-Mercer smoothing*

Add WeChat powcoder

Assignment Project Exam Help

- ▶ Difference between Web search and Information Retrieval.
- ▶ Estimation of relative sizes of two search engines.
- ▶ Near duplicate detection: the shingling method
- ▶ **Not required:** the SimHash method.

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

- ▶ Understand the requirements and the current architecture of crawlers (e.g., the Mercator architecture).
- ▶ **Not required:** optimization for age, finding content blocks, etc.

Add WeChat powcoder

Assignment Project Exam Help

- ▶ The pagerank algorithm: theory and practice
- ▶ Not required: the topic-specific/personalized pagerank

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Thanks and Good Luck!
Add WeChat powcoder