



1. Web Mining can be classified into three categories. What are they? (1.5 pts)

يمكن تصنيف تنقيب الويب إلى ثلاث فئات، ما هي؟

2. Cite four tasks among the most commonly used tasks in data mining, and give one example for each task. (4 pts)

اذكر أربع مهام من أكثر المهام استخدامًا في تنقيب البيانات، واذكر مثالًا لكل مهمة

3. Why is TF-IDF weighting more favorable for representing a document as a vector compared to other representations (Boolean, term frequency)? (2 pts)

لماذا تُعدّ أوزان TF-IDF أكثر ملاءمة لتمثيل الوثيقة على شكل شعاع مقارنة بالتمثيلات الأخرى (التمثيل المنطقي، تكرار المصطلحات)

4. What are hub and authority pages, and in which algorithm were they introduced? (2pts)

ما هي صفحات المحاور (Hubs) وصفحات السلطات (Authorities)، وفي أي خوارزمية تم تقديمهما؟

5. What is Latent Semantic Indexing (LSI), and what problem does it aim to solve? (1.5 pts)

ما هو الفهرسة الدلالية الكامنة (LSI)، وما المشكلة التي تهدف إلى حلّها؟

6. What is a confusion matrix, and how is it used to evaluate classification models? (1.5 pts)

ما هي مصفوفة الالتباس (Confusion Matrix)، وكيف تُستخدم في تقييم نماذج التصنيف؟

7. What are web communities, and why is discovering them useful for search engines? (2pts)

ما هي مجتمعات الويب، ولماذا يُعد اكتشافها مفيدًا لمحركات البحث؟

8. What are the strengths and weaknesses of the PageRank algorithm in evaluating the popularity of web pages? (2pts)

ما هي نقاط القوة ونقاط الضعف في خوارزمية PageRank عند تقييم شعبية صفحات الويب؟

9. Classify these (presented) algorithms into two categories: supervised learning and unsupervised learning. (3.5 pts)

صنّف هذه الخوارزميات (المعروضة) إلى فئتين: التعلم المُراقب والتعلّم غير المُراقب.

- | | | |
|---------------------------|-------------------------|---------------|
| • Gaussian Mixture Models | • FCM. | • SVM |
| • Decision trees. | • Fp-Growth. | • PCA |
| • DBSCAN | • Hierarchical methods. | • Naïve Bayes |
| | • Neural networks. | |



Level: 2nd year Master RTIC

Subject: Web Mining

Exam

Academic year: 2025/2026

Duration: 2h :0m

1. Web Mining categories:
 - Web Content Mining
 - Web Structure Mining
 - Web Usage Mining
2. Common data mining tasks (with examples):
 - Classification: spam detection
 - Regression: house price prediction
 - Clustering: user segmentation
 - Association rules: market basket analysis
3. TF-IDF:
 - It highlights important terms and reduces the impact of common words.
4. Hub and authority pages:
 - Authority: many incoming links
 - Hub: many outgoing links
 - Algorithm: HITS
5. LSI:
 - A dimensionality reduction method that handles synonymy and polysemy.
6. Confusion matrix:
 - A table used to evaluate classification performance (accuracy, precision, recall).
7. Web communities:
 - Groups of closely linked pages; useful for improving search results.
8. PageRank:
 - Strengths: measures global importance using links
 - Weaknesses: query-independent, favors old pages
9. Algorithms:
 - Supervised: Decision Trees, Neural Networks, SVM, Naïve Bayes
 - Unsupervised: GMM, DBSCAN, FCM, FP-Growth, Hierarchical, PCA