University of Coimbra

SEMESTER I 2025-2026

Knowledge and Language

January 2026 TIME ALLOWED: 2 HOURS

DISCLAMER

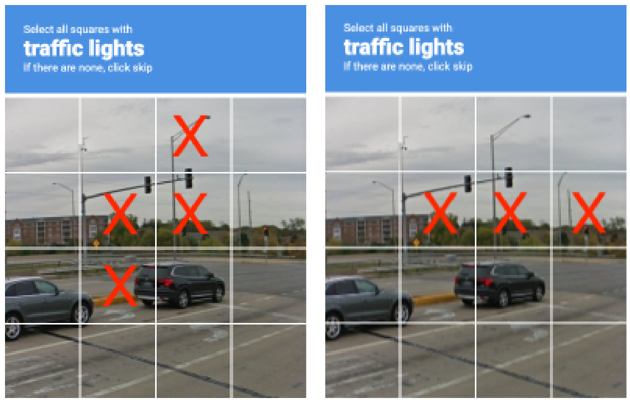
1. This exam is **NOT** oficial nor mimics the structure of the real exam.
2. Any typos or mistakes please inform tds@student.uc.pt
3. Based on course slides and theoretical classes

DOCUMENT ATTRIBUTION

This exam document was authored and compiled by **Tiago Silva** (2022216215). It is not an official examination issued by the University of Coimbra.

1. Human-AI interaction

Question 1.1 (Cohen Kappa) Find the level of agreement between the two annotators.



Question 1.2 Within the scope of Human-Centered AI, what are the various roles humans can play, ranging from supervisors and annotators to validators and regulators

Question 1.3 How do model-agnostic post-hoc explainability methods (such as LIME or SHAP) allow for the interpretation of "black-box" model decisions without altering their original architecture

Question 1.4 How do data annotation and crowdsourcing (Wisdom of Crowds) influence the quality and consistency of AI systems, and how is this reliability measured using the Kappa coefficient

Author: Tiago Silva. Based on course slides by Hugo Gonçalo Oliveira. Not an official exam of the University of Coimbra.

2. Knowledge Representation

Question 2.1 (Knowledge Representation) Represent the following sentences in first-order logic, using a consistent vocabulary.

- Some students took MAS in 2019.
- Every student who takes KL passes it.
- In 2024, there was a course taken by a single student.
- The best score in KL is always higher than the best score in MAS.

Question 2.2 (Knowledge Representation) Create a KG that contains the following classes:

- Computer
- Hard Drive
- Laptop
- Energy
- Battery
- Screen
- Pixel

Question 2.3 What are the primary differences between Propositional Logic and First-Order Logic (FOL) regarding their ontological commitments and their capacity to represent objects and relations generically

Question 2.4 How do frame-based representations (object vs. event frames) allow for the organization of stereotypical knowledge into hierarchical structures that can fill informational gaps?

Question 2.5 How do ontologies formalize the intended meaning of a vocabulary, and how do OWL constraints (e.g., allValuesFrom, someValuesFrom) enable automated inference?

3.Vocabularies

Question 3.1 Interpret the following RDFs:

```xml
<rdfs:Class rdf:ID="#FlashMemory">
  <rdfs:subClassOf>
    <rdfs:Class rdf:ID="#Consumable"/>
  </rdfs:subClassOf>
</rdfs:Class>
<rdfs:Class rdf:ID="#Printer">
  <rdfs:subClassOf>
    <rdfs:Class rdf:ID="#Hardware"/>
  </rdfs:subClassOf>
</rdfs:Class>
<rdfs:Class rdf:ID="#LCD">
  <rdfs:subClassOf>
    <rdfs:Class rdf:ID="#Monitor"/>
  </rdfs:subClassOf>
</rdfs:Class>
```

```xml
<rdf:Property rdf:ID="#Capacity">
  <rdfs:domain rdf:resource="#FlashMemory">
  <rdfs:range rdf:resource="xsd:int"/>
</rdf:Property>
<rdf:Property rdf:ID="#PagesPerMinute">
  <rdfs:domain rdf:resource="#Printer">
  <rdfs:range rdf:resource="xsd:int"/>
</rdf:Property>
<rdf:Property rdf:ID="#Model">
  <rdfs:domain rdf:resource="#Hardware">
  <rdfs:domain rdf:resource="#Consumable">
  <rdfs:domain rdf:resource="#Monitor">
  <rdfs:range rdf:resource="xsd:string"/>
</rdf:Property>
```

Question 3.2 Express the following sentences using RDF/OWL schemas:

- "An animal eats living things"
- "A parent of a living thing is a living thing"
- "A parent is an animal and parent of animals"
- "A pet is an animal with a single human owner"

Question 3.3 Express, in natural language, the following RDF/OWL schema:

```xml
<owl:Class rdf:ID="UC_Engineer">
  <rdfs:subClassOf rdf:resource="#Engineer"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasFriend"/>
      <owl:someValuesFrom rdf:resource="#Engineer"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasQuality"/>
      <owl:allValuesFrom rdf:resource="#GoodQuality"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#studiedAt"/>
      <owl:hasValue="#UC"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

4. Semantic Web and Linked Data

Question 4.1 Using the RDF schema:

```
@prefix ex: <http://example.org/> .
@prefix taubz: <http://razor.occams.info/index.html#> .
taubz:me            ex:own      taubz:my_apartment .
taubz:me            ex:own      taubz:my_computer .
taubz:my_apartment  ex:contains taubz:my_computer .
taubz:my_apartment  ex:contains taubz:friends_junk .
taubz:my_apartment  ex:location <http://example.org/Philadelphia> .
taubz:me            ex:own      taubz:my_desk .
taubz:my_desk       ex:contains taubz:my_pens_and_pencils .
```

What are the results of the following query:

```
PREFIX ex: <http://example.org/>
PREFIX taubz: <http://razor.occams.info/index.html#>
SELECT ?what
WHERE {
    taubz:me ex:own ?container .
    ?container ex:contains ?what .
}
```

Question 4.2 What are the four fundamental rules of Linked Data proposed by Tim Berners-Lee, and how does the use of HTTP URIs guarantee information accessibility?

Question 4.3 How does the RDF model (subject-predicate-object) and its various serialization formats (XML, Turtle, JSON-LD) enable the creation of a global, interconnected knowledge graph?

Question 4.4 In what way do the FAIR principles (Findable, Accessible, Interoperable, Reusable) guide the publication of linked data to ensure it is processable by machines?

Question 4.5 Write the SPARQL queries (extract from DBPedia):

- Works authored by someone named "Bob Dylan"
- Works authored by winners of the Nobel Prize in Literature
- English names of winners of the Nobel Prize in Literature and of works authored by them, with the language and genre of the work, if available

5. Uncertainty and Bayesian Networks

Question 5.1 Use Naive Bayes for predicting whether a new customer (aged 28, medium income, student with a fair credit rating) will buy a computer.

Row	Age	Income	Student	Credit Rating	Buys Computer?
1	$\leq$30	high	no	fair	no
2	$\leq$30	high	no	excellent	no
3	31–40	high	no	fair	yes
4	$>$40	medium	no	fair	yes
5	$>$40	low	yes	fair	yes
6	$>$40	low	yes	excellent	no
7	31–40	low	yes	excellent	yes
8	$\leq$30	medium	no	fair	no
9	$\leq$30	low	yes	fair	yes
10	$>$40	medium	yes	fair	yes
11	$\leq$30	medium	yes	excellent	yes
12	31–40	medium	no	excellent	yes
13	31–40	high	yes	fair	yes
14	$>$40	medium	no	excellent	no

Question 5.2

> *"You want to diagnose whether there is a fire in a building. You receive a noisy report about whether everyone is leaving the building. If everyone is leaving, this may have been caused by a fire alarm. If there is a fire alarm, it may have been caused by a fire or by tampering. If there is a fire, there may be smoke."*

- Build the Bayesian Network for this problem.
- How many parameters do you need to represent the Full Joint Probability Distribution? What if using a Bayes Net?
- What is the probability that the alarm is tampered, there is fire, the alarm rings, there is no smoke and people are leaving the building?

Question 5.3 How does the graphical structure of a Bayesian Network, specifically its directed acyclic graph (DAG) topology, allow for a compact representation of full joint distributions through assertions of conditional independence?

Question 5.4 In what way does local semantics—where each node is independent of its non-descendants given its parents—relate to global semantics to ensure the full joint probability distribution can be reconstructed?

Question 5.5 What is the specific role of the Markov blanket in determining the boundaries of a node's independence relative to all other variables in the network?

Question 5.6 Why does the number of parameters required for a Boolean network with n variables grow linearly (O($n \cdot 2^k$)) rather than exponentially (O(2^n)) when the maximum number of parents k is limited?

Question 5.7 How does the initial ordering of variables during the construction of a network influence the resulting sparsity of its structure and the ease of deciding conditional independence?

Question 5.8 To what extent can the introduction of hidden variables or functional relationships (such as noisy-OR) reduce the total number of parameters needed in the Conditional Probability Tables (CPTs) of a diagnostic system?

Question 6.1 A machine can be in one of two states: **Working (W)** or **Faulty (F)**. The true state cannot be observed directly. Instead, a signal light is observed: **Green (G)** or **Red (R)**.

The **state transition probabilities** are:

$$\begin{bmatrix} P(W_{t+1}|W_t) & P(F_{t+1}|W_t) \\ P(W_{t+1}|F_t) & P(F_{t+1}|F_t) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$$

The **emission probabilities** are:

$$\begin{bmatrix} P(G|W) & P(R|W) \\ P(G|F) & P(R|F) \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

Given:

- Today's observation: Red light (R)
- Prior belief: $P(W) = 0.5$, $P(F) = 0.5$

Tasks:

1. Compute the posterior probability of the machine being in each state today.
2. Predict the probability of the machine being in each state tomorrow.

Question 6.2 (HMM) A person's mood can be in one of two hidden states: **Happy (H)** or **Sad (S)**. The mood cannot be observed directly. Instead, we observe the person's behavior: **Talkative (T)** or **Quiet (Q)**.

The **state transition probabilities** are:

$$\begin{bmatrix} P(H_{t+1}|H_t) & P(S_{t+1}|H_t) \\ P(H_{t+1}|S_t) & P(S_{t+1}|S_t) \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

The **emission probabilities** are:

$$\begin{bmatrix} P(T|H) & P(Q|H) \\ P(T|S) & P(Q|S) \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}$$

Given:

- Today's observation: Talkative (T)

6. Probabilistic Reasoning Overtime

- Prior belief: $P(H) = 0.5$, $P(S) = 0.5$

Tasks:

1. Compute the posterior probability of the mood today using the observation (filtering).
2. Predict the probability of each mood tomorrow using the posterior as the prior.

Question 6.3 What is the fundamental theoretical and practical distinction between first-order and second-order Markov processes, and how does the sensor Markov assumption simplify the relationship between evidence and unobservable states?

Question 6.4 How do the four primary temporal inference tasks—filtering, prediction, smoothing, and most likely explanation—differ in terms of the evidence used and their specific statistical objectives?

Question 6.5 In what way does the Viterbi algorithm differ from conventional filtering when seeking the sequence of states that best explains an entire sequence of observations ?

7. Text Classification & Knowledge Extraction

Question 7.1 Classify the sentence "tanta espera, mas fraca qualidade" using Naive Bayes

Training Data:

Class	Content
−	experiência muito fraca
−	grande desilusão
−	muito tempo à espera
+	muito boa relação qualidade preço
+	excelente serviço

Question 7.2 Which one is the best classifier?

There are 200 comments about a bakery available on the Web:

- 130 are positive (4 or 5 stars)
- 70 are not positive (1, 2, or 3 stars)

Two models are available for polarity detection:

1. **Naive Bayes (NB)**: classified 150 comments as positive, out of which 110 were actually positive.
2. **Rule-Based (RB)**: classified 100 comments as positive, out of which 80 were actually positive.

Confusion Matrices:

Naive Bayes (NB):

NB Classified	Positive	Negative
Positive	110	40
Negative	20	30

Rule-Based (RB):

RB Classified	Positive	Negative
Positive	80	20
Negative	50	50

Question 7.3 PoS tagging with HMMs Consider a limited language with only three part-of-speech tags:

- D: Determiners, $P(D) = 0.3$
- N: Nouns, $P(N) = 0.4$
- V: Verbs, $P(V) = 0.3$

Sequence start symbol: S.
Transition Probabilities:

$P(\cdot\|\cdot)$	D	N	V
S	0.4	0.4	0.2
D	0	1	0
N	0.2	0.1	0.7
V	0.6	0.2	0.2

Emission Probabilities:

$P(\text{word}\|\text{tag})$	the	a	water	talk	hear
D	1	0.9	–	–	–
N	–	0.1	0.7	0.2	–
V	–	–	0.3	0.8	1

Task:
For the following sequences:

1. the water
2. a talk
3. hear a talk

- Determine the best tag sequence for each sentence based on the HMM.
- Use the Viterbi algorithm to find the most probable tag sequence.
- Suggestion: draw the graphical model and consider all possible paths for the tag sequences.

Question 7.4 WSD with Sense Embeddings Which sense would bank be mapped to?

Consider the word *bank* with the following sense embeddings:

Sense	Example Sentences	Sense Embedding
bank.1	he cashed a check at the bank; that bank holds the mortgage on my home	`[0.22, 0.12, 0.37, 0.27, 0.12]`
bank.2	they pulled the canoe up on the bank; he sat on the bank of the river and watched the currents	`[0.07, 0.32, 0.12, 0.42, 0.22]`
bank.3	a huge bank of earth	`[0.08, 0.28, 0.16, 0.38, 0.19]`
bank.4	he operated a bank of switches	`[0.18, 0.22, 0.14, 0.32, 0.20]`

Given the context embedding for *bank*:

$$c = [0.23, 0.12, 0.38, 0.28, 0.14]$$

Question 7.5 Within the context of knowledge extraction, what is the significance of Sequence Labeling for critical tasks such as Named Entity Recognition (NER) and Part-of-Speech (PoS) tagging

Question 7.6 How does the technique of distant supervision leverage large external databases to automatically label massive volumes of text for relation extraction

Question 7.7 In what way does Open Information Extraction (Open IE) allow for the retrieval of relational triples without the need to specify predefined relation types beforehand?

Author: Tiago Silva. Based on course slides by Hugo Gonçalo Oliveira. Not an official exam of the University of Coimbra.

8. Sentiment Analysis

Question 8.1 In what way does Open Information Extraction (Open IE) allow for the retrieval of relational triples without the need to specify predefined relation types beforehand?

Question 8.2 What are the independence assumptions of the Naive Bayes model, and how can TF-IDF weighting be incorporated to improve the classification of text documents?

Question 8.3 Which primary linguistic challenges, such as negation, anaphora, and figurative language (irony), limit the effectiveness of approaches based solely on sentiment lexicons?

9. NLP, Distributional Semantics & Word Embeddings

Question 9.1 What unique challenges do phenomena such as ambiguity (phonological, lexical, syntactic, and semantic) and vagueness impose on natural language processing compared to formal programming languages?

Question 9.2 How does the traditional NLP pipeline (morphological, syntactic, and semantic analysis) compare to modern approaches based on Zero-shot LLM Prompting?

Question 9.3 How are conversational agents distinguished based on their primary goals—such as task-oriented Virtual Assistants versus chat-oriented Companions—and what were the core limitations of the early ELIZA model in this evolution?

Question 9.4 How does the distributional hypothesis, which suggests a word is characterized by the "company it keeps," serve as the foundation for creating term co-occurrence matrices?

Question 9.5 What are the primary advantages of using dense, lower-dimension word embeddings (such as Word2Vec or GloVe) over sparse term-count matrices, and how are linguistic regularities maintained in these spaces?

Question 9.6 How do contextualized word embeddings (based on Transformers) resolve the limitations of static embeddings when dealing with polysemous terms like "bank"?

10. Large Language Models & Knowledge

Question 10.1 What characterizes the paradox that LLMs can generate fluent and coherent text without possessing a real understanding of language, often being described as "Stochastic Parrots"?

Question 10.2 How do techniques like Retrieval Augmented Generation (RAG) and Knowledge Injection mitigate the problem of factual hallucinations in large language models?

Question 10.3 Why is the distinction between plausible explanations (human-coherent) and faithful explanations (representative of the model's internal reasoning) critical for the analysis of LLMs?

END OF PAPER

Author: Tiago Silva. Based on course slides by Hugo Gonçalo Oliveira. Not an official exam of the University of Coimbra.