

Quiz 1 RUBRICS

1. Which of the following is NOT a named entity in the sentence: "Dr. Sarah visited the Johns Hopkins Hospital in Baltimore for a medical conference"?
 - a. Dr. Sarah
 - b. Johns Hopkins Hospital
 - c. **visited**
 - d. Medical conference

2. What does FHIR stand for?
 - a. Findable Healthcare Interactive Reusable
 - b. Fair Healthcare Interoperability Resources
 - c. **Fast Healthcare Interoperability Resources**
 - d. Findable Healthcare Interoperable Reusable

3. Which of the following is NOT a purpose of Electronic Health Records (EHR)?
 - a. Access to health records of a patient
 - b. Seamless integration of Billing and Insurance
 - c. **Secure messaging between patients**
 - d. Improvement of Quality of Care

4. Which of the following is NOT true about HL7 messages?
 - a. They are delimited by a pipe separator
 - b. They can have multiple segments per message
 - c. They are used primarily for medical data exchange
 - d. **They are restricted to text-only formats**

5. Which type of model describes how an EHR is conceptually designed in a system, independent of technology?
 - a. Platform-Specific Model
 - b. **Computational-Independent Model**
 - c. Platform-Independent Model
 - d. Wire Format Model

6. In the context of interoperability, which of the following refers to preserving meaning during the exchange of information?
 - a. **Semantic Interoperability**
 - b. Technical Interoperability
 - c. Human Interoperability
 - d. Syntactic Interoperability

7. What is the major difference between SNOMED CT's primitive and fully defined concepts?
 - a. Primitive concepts have unique relationships distinguishing them

- b. Fully defined concepts lack defining relationships
 - c. Primitive concepts do not have unique relationships that distinguish them
 - d. Fully defined concepts cannot be reused
8. Which of the following is NOT mandatory in a FHIR resource?
- a. Human-readable summary
 - b. Identifier
 - c. URL link
 - d. Profiling
9. In a clinical setting, what is the primary purpose of a Refined Message Information Model (RMIM) in HL7?
- a. Provide raw data for machine learning models
 - b. Refine the abstract concepts of RIM for specific use cases
 - c. Generalize data definitions for international use
 - d. Design a medical device

Short-Answer Questions (ATTEMPT ANY 5)

5x5 = 25

1. What are the three main levels of interoperability? Provide one example for each.

Ans. The three main levels of interoperability are:(any three)

- Technical Interoperability
- Syntactic Interoperability
- Semantic Interoperability
- Process
- Human

1. Technical Interoperability

Definition: Refers to the ability of different systems to exchange data at a basic level, focusing on the technical process of data transfer. It ensures that systems can communicate and exchange data reliably.

Example: The Health Level 7 (HL7) messaging standard ensures that systems like electronic health records (EHRs) and laboratory information systems (LIS) can exchange information via a shared technical protocol.

2. Syntactic Interoperability

Definition: Ensures that the structure or format of the exchanged data is understandable between systems. It standardizes the format and structure of the data being exchanged.

Example: Fast Healthcare Interoperability Resources (FHIR) provides a syntactic standard, ensuring that healthcare data follows a common format (e.g., JSON or XML) when shared across systems.

3. Semantic Interoperability

Definition: Refers to the ability of systems to interpret and use the exchanged data meaningfully. It ensures that the meaning of the data remains consistent and accurate across different systems.

Example: SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) ensures semantic interoperability by standardizing clinical terminology, enabling systems to interpret data in the same way.

4. Process Interoperability

Definition: Focuses on aligning business processes and workflows between different systems to enable collaboration. This level of interoperability ensures that systems not only exchange data but also that the processes are synchronized.

Example: In a hospital, a lab system automatically notifying a physician's EHR system about a test result and triggering a workflow for follow-up treatment reflects process interoperability.

5. Human Interoperability is sometimes considered an additional level of interoperability, focusing on the ability of individuals (clinicians, healthcare providers, etc.) to effectively interpret and use the data exchanged between systems. This level addresses the human aspect of interaction with interoperable systems, emphasizing communication, collaboration, and decision-making based on shared data.

Example: When a physician receives lab results through an EHR system, they must interpret the data correctly and decide on the next course of action (e.g., medication adjustments). Human interoperability ensures that healthcare professionals can collaborate effectively, communicate findings clearly, and make informed decisions based on shared data from different systems.

2. Explain how Syntactic Interoperability is achieved using HL7 and FHIR standards. (lecture3)

Ans.

Criteria	Marks	Description
----------	-------	-------------

Definition of Syntactic Interoperability	1	Ensures that the structure or format of the exchanged data is understandable between systems. It standardizes the format and structure of the data being exchanged.
Introduction to HL7	1	HL7 (Health Level 7) is an international set of standards designed for the exchange, integration, sharing, and retrieval of electronic health information. It plays a critical role in achieving syntactic interoperability by providing structured messaging formats
Explanation of HL7 in achieving Syntactic Interoperability	1	HL7 v2 is widely used to structure healthcare data such as patient admissions, lab results, and medical records in a consistent, standardized format that different systems can interpret. By ensuring that all systems follow the same message structure and formatting rules, HL7 allows for the seamless exchange of data between disparate healthcare systems, achieving syntactic interoperability.
Introduction to FHIR	1	FHIR (Fast Healthcare Interoperability Resources) is a newer standard developed by HL7 that builds on previous standards but focuses on modern web technologies and simpler, more flexible data structures
Explanation of FHIR in achieving Syntactic Interoperability	1	FHIR enables syntactic interoperability by defining standardized Resources (like Patient, Observation, and Medication), which serve as building blocks for healthcare data. These Resources use consistent data models and formats (such as JSON or XML), ensuring that data exchanged between systems is well-structured and easily parsed. FHIR's design principles, including its use of modern web protocols like REST, further simplify the process of exchanging structured data across healthcare systems.

3. What is the role of SNOMED CT in medical ontologies, and how does it help in clinical documentation? (lecture 4&5)

Ans.

Criteria	Marks	Description
Definition of SNOMED CT	1	SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) is a comprehensive, multilingual clinical healthcare terminology that provides a structured way to encode medical terms.
Role of SNOMED CT in medical ontologies	1	It is an essential part of medical ontologies because it organizes clinical concepts in a hierarchical structure, allowing for the accurate classification of diseases, treatments, and other health-related concepts
How SNOMED CT supports clinical documentation	2	SNOMED CT plays a crucial role in clinical documentation by enabling standardized coding of medical records, which ensures consistent and precise documentation across healthcare providers. This uniformity supports electronic health records (EHR) systems in capturing detailed clinical information, improving the quality of data and reducing ambiguities in medical terminology.
Interoperability and data sharing benefits	1	SNOMED CT enhances interoperability , as it enables different healthcare systems to communicate using a common language. This standardized approach also helps in clinical decision-making, research, and data analytics by ensuring that information is represented consistently across systems.

4. Explain the concept of a Bag of Words (BoW) model. What are two main challenges associated with it?

Ans.

Criteria	Marks	Description
Explanation of the Bag of Words model	2	The Bag of Words (BoW) model is a simple and widely used technique in natural language processing (NLP) that represents text by treating it as a collection of independent

		<p>words, ignoring grammar and word order.</p> <ul style="list-style-type: none"> - Assumes that the text belonging to a given class in the dataset is characterized by a unique set of words - Knowing the words present in a text tells about the class (bag) - Each document is a V-dimensional vector - It gives a fixed-length representation - "Dog Bites Man" = [1 1 1 0 0 0]
First Challenge	1	<p>Any 2 of the following:</p> <p>- Loss of Context Since BoW treats words as independent entities, it fails to capture the meaning or relationship between words. For example, the phrases "dog bites man" and "man bites dog" would be represented identically, despite their very different meanings.</p>
Second Challenge	1	<p>-High Dimensionality Since the model relies on a large vocabulary, it leads to sparse vectors, especially when dealing with large corpora, as many words in the vocabulary will not appear in a given document. This makes the model computationally expensive and harder to manage.</p> <p>- Ignoring Semantic Meaning BoW does not consider the semantic meaning or relationships between words. For instance, it treats synonyms as distinct features, even though they might have similar meanings. This limitation can lead to a loss of information that could otherwise help in understanding the document's context.</p> <p>- Inability to Capture Word Order: BoW disregards the order of words, meaning it treats "The cat sat on the mat" and "The mat sat on the cat" as identical. This can lead to confusion in cases where word order is essential for understanding the meaning.</p> <p>-Handling of Stop Words: Without proper handling, common words like "the," "is," and "and" (stop words) may dominate the feature space, reducing the effectiveness of the model. While this can be mitigated by removing stop words, it can sometimes lead to loss of important context.</p> <p>-Scalability Issues with Large Corpora: For large corpora, the size of the vocabulary can grow exponentially, leading to scalability problems in both computational cost and memory usage. This makes BoW difficult to use for very large datasets without efficient dimensionality reduction techniques.</p>
Example or elaboration	1	

5. What is the key difference between One-Hot Encoding and Neural Word Embeddings?

The key difference between One-Hot Encoding and Neural Word Embeddings lies in the way they represent data and the information they capture.

Representation (2 Marks):

- **One-Hot Encoding** represents each word as a sparse binary vector, where only one element is 1 (indicating the presence of the word), and all other elements are 0. Each word is assigned an arbitrary, independent vector with no inherent relationships between words.
- **Neural Word Embeddings (e.g., Word2Vec, GloVe)** represent words as dense, continuous vectors in a lower-dimensional space. These vectors capture semantic relationships between words, so words with similar meanings or contexts are closer together in this space.

Dimensionality and Efficiency (1 Mark):

- **One-Hot Encoding** has high dimensionality (equal to the vocabulary size), making it computationally inefficient and memory-intensive for large vocabularies.
- **Neural Word Embeddings** have significantly lower dimensions, making them more efficient in terms of both computation and memory usage.

Semantic Information (2 Marks):

- **One-Hot Encoding** does not capture any semantic or contextual information; each word is treated as completely independent.
- **Neural Word Embeddings** capture meaningful relationships and contexts between words. For example, in embeddings, words like "king" and "queen" will have similar vectors that reflect their related meanings.

6. What is the difference between the skip-gram and continuous Bag of Words (CBOW) in Word2Vec model?

Ans. Core Objective :

Skip-gram: Predicts the context words (surrounding words) given a target word.

CBOW: Predicts the target word based on the context (surrounding words).

Data Representation:

Skip-gram: Uses the target word as input and tries to predict the nearby context words.

CBOW: Uses multiple context words as input and tries to predict the target word in the center.

Training Efficiency:

Skip-gram: Slower training, especially when the context window is large, as it predicts multiple output words for each input word.

CBOW: Faster training because it predicts only one word based on the combined context.

Performance with Rare Words:

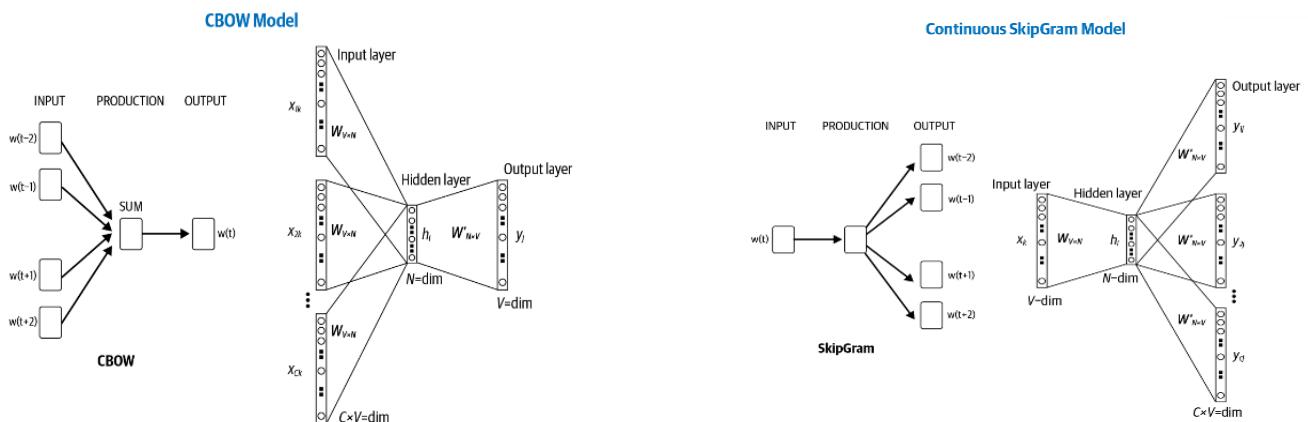
Skip-gram: Performs better for infrequent words because it learns specific context from the target word.

CBOW: Struggles with rare words as it focuses on predicting common words from aggregated context.

Use Cases:

Skip-gram: Useful when the corpus contains rare words and phrases.

CBOW: More suitable for faster general-purpose training on large datasets with frequent words.



Long Answer Questions (ATTEMPT ANY 2)

10x2 = 20

1. What are Ontologies? Explain the difference between Concepts, Instances and Data using an example from healthcare. What property of the Ontology graph makes computational Decision Support possible?

Ans.

Criteria	Marks	Description
Definition of Ontologies	2	Ontologies are formal representations of knowledge within a specific domain. They define concepts, their attributes, and relationships between them, enabling structured and meaningful data exchange. In healthcare, ontologies represent medical knowledge, such as diseases, treatments, and patient information.
Explanation of Concepts, Instances, and Data	4	In an ontology, Concepts represent abstract or general categories. For example, "Disease" is a concept. Instances are specific realizations of a concept. For example, "Diabetes" is an instance of the "Disease" concept. Data refers to specific, concrete information. For example, a patient's blood glucose level of 150 mg/dL is data associated with the instance "Diabetes." Example from Healthcare: In a healthcare ontology, "Disease" is a concept, "Diabetes" is an instance of that concept, and a patient's blood sugar levels or treatment plans are examples of associated data. These ontologies allow standardization and ensure accurate information sharing between healthcare systems
Ontology Graph Property for Decision Support	2	The property of the Ontology graph that makes computational decision support possible is its ability to infer new relationships based on existing ones. The structured relationships between concepts and instances allow systems to draw logical conclusions, making the ontology useful for clinical decision support (e.g., identifying treatment plans based on a patient's diagnosis). The directed and acyclic nature of an ontology graph is also an important property that enhances computational decision support. Directed edges ensure that relationships between concepts (such as "is-a," "part-of," or "causes") have a clear direction, which is essential for logical reasoning. For instance, knowing that "Diabetes is a Disease" and not the reverse helps the system apply logical inferences correctly.

		Acyclicity prevents cycles in the graph, which means that no concept refers back to itself in a loop. This ensures that reasoning or traversal through the graph terminates and avoids infinite loops, making computations more efficient and reliable.
--	--	---

2. **What is a vector space model(VSM)? Explain how compositional algebra works in a Vector Space Model. Taking an example such as "king - man + woman = queen", give the mathematical basis and a vector diagram to illustrate your answer.**

Ans.

Criteria	Marks	Description
Definition of Vector Space Model (VSM)	2	Clear definition of VSM, mentioning its use in representing words as vectors in a high-dimensional space.
Explanation of Compositional Algebra	3	Detailed explanation of how compositional algebra works in VSM, involving vector arithmetic (addition, subtraction) to capture relationships.
Mathematical Explanation (King - Man + Woman)	3	Demonstrates understanding of the algebraic manipulation of vectors (e.g., king - man + woman = queen) with a mathematical basis.
Vector Diagram	2	Clear and correct illustration of the vector diagram showing word relationships in vector space.

Definition of VSM (2 marks):

A **Vector Space Model (VSM)** represents text data as vectors in a multi-dimensional space. Each word or document is assigned a vector, and the similarity between texts is measured using the distance or angle between their vectors. VSM is commonly used in information retrieval and NLP for tasks like document comparison or word similarity.

Explanation of Compositional Algebra (3 marks):

Compositional algebra in a VSM refers to combining or manipulating vectors to produce new

ones based on relationships between words. For example, if you take the vector for "king", subtract the vector for "man", and add the vector for "woman", the result approximates the vector for "queen". This approach models the semantic relationships between words based on their vector distances and directions.

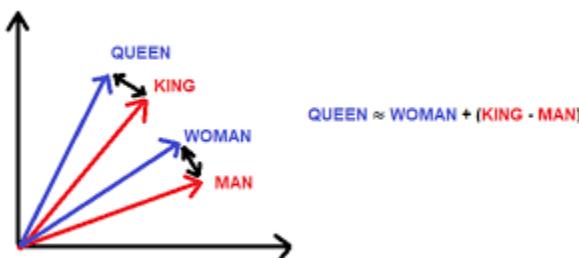
Mathematical Explanation with Example (3 marks):

Using vectors, the equation " $\text{king} - \text{man} + \text{woman} = \text{queen}$ " means that the vector difference between "king" and "man" captures the concept of royalty without gender. Adding "woman" shifts the vector towards female royalty, resulting in "queen". Mathematically:

$$\vec{\text{king}} - \vec{\text{man}} + \vec{\text{woman}} \approx \vec{\text{queen}}$$

The difference vectors $(\vec{\text{king}} - \vec{\text{man}})$ and $(\vec{\text{queen}} - \vec{\text{woman}})$ are expected to be similar in the vector space.

Vector Diagram Explanation (2 marks):



$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

A vector diagram would show vectors for "king", "man", "woman", and "queen" positioned in a multi-dimensional space. The vectors for "man" and "woman" would point in opposite directions from their corresponding gender-neutral royal terms, while the vectors for "king" and "queen" would align closely. This illustrates the algebraic relationships between the words.

3. Design a healthcare monitoring system for diabetes chronic disease management. Create a conceptual process model (hint: similar to a UML activity diagram) that highlights patient interactions with the system, physician inputs, and decision support tools. How will you implement semantic interoperability within such a system?

Ans.

Criteria	Marks	Description
Description of System Design	2	Clear description of the healthcare monitoring

		system, focusing on diabetes management.
Explanation of Key Components	2	Detailed explanation of the system's components: patient monitoring, physician inputs, and decision support tools.
Conceptual Process Model	3	Well-structured conceptual process model (UML activity diagram or similar), showing the flow of interactions between patients, physicians, and the system.
Implementation of Semantic Interoperability	3	Explanation of how semantic interoperability will be achieved, referencing standards (e.g., SNOMED CT, HL7 FHIR) and methods for ensuring data compatibility.

System Overview (2 marks):

A healthcare monitoring system for diabetes management would involve regular tracking of patient vitals (e.g., blood sugar levels, insulin doses, diet, etc.) and communication with healthcare providers. The system could include patient devices (e.g., glucose meters), a mobile app for tracking and reminders, and physician dashboards for monitoring patient health.

Conceptual Process Model (Activity Diagram) (4 marks):

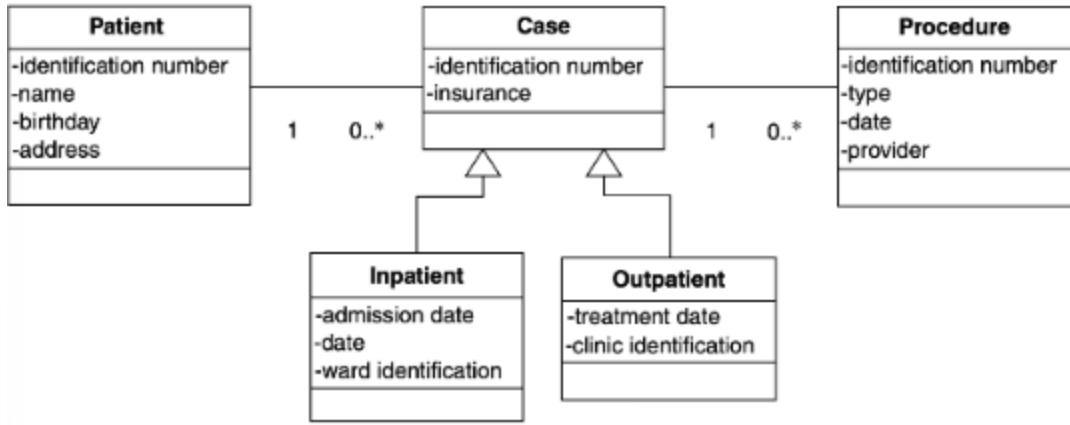
The conceptual process model should include the following steps:

Patient Interaction: The patient logs daily glucose readings, meal information, and medication doses through the system (via a mobile app or wearable device).

Physician Input: Physicians access patient data to review trends, analyze data, and adjust treatments as necessary. They may also interact with a decision support tool that suggests treatment plans or flags potential complications.

Decision Support Tool: An AI-powered tool reviews patient data and provides recommendations based on medical guidelines. It could suggest insulin dosage adjustments or flag abnormal trends in glucose levels for physician review.

Diagrammatically, these interactions would be represented as actions (patient logs data, physician reviews, etc.) and decisions (e.g., adjustments to treatment plans based on data trends).



(Based on your system)

Semantic Interoperability (4 marks):

Semantic interoperability ensures that data exchanged between devices, applications, and healthcare providers is meaningful and consistent. In this system, it would be implemented using standards such as FHIR (Fast Healthcare Interoperability Resources) and SNOMED CT. For example, glucose readings from different devices should be interpreted in the same way, regardless of device manufacturer. By adhering to standardized ontologies and vocabularies, the system allows for seamless data integration and shared understanding across various platforms and users.