Computing For Medicine
CSE585/BIO546
Mid Semester Examination

Maximum Marks: 50 **Duration : 60 minutes**

SECTION -A [MULTIPLE CHOICE QUESTIONS] -5 marks (1 mark x 5)

1. Which of the following FHIR resources is useful for checking the validity of a FHIR implementation?
   a. Conformance
   b. Medications
   c. Organization
   d. Patient

2. _____ model predicts the probability of the target word given its context defined within a window.
   a. Continuous Bag of Words
   b. Skip Gram model
   c. Bag of words
   d. N-gram model

3. Inducing a "low dimensional embedding", i.e. one that is NOT sparse, can be achieved:

   a) Tokenization
   b) Distributed Representations
   c) Stemming
   d) One-Hot Encoding

4. Which of the following is NOT true with respect to One-Hot Encoding:
   a) Storage constraints and overfitting due to sparsity
   b) Does not give fixed length representation
   c) Does not assume independence between words
   d) Out of Vocabulary problem- needs retraining every time a new word is added

5. Which of the following is a Character Level word embedding
   a) GloVe
   b) Fasttext
   c) Word2Vec
   d) None of the above

SECTION-B [SHORT ANSWER QUESTIONS] - 21 marks (3 marks x 7)

PLEASE EXPLAIN IN THREE SENTENCES EACH

a) **Directed acyclic graphs?**
A graph with directed edges between nodes. There are no cycles in a DAG. Commonly used applications of DAGs include Ontologies and reasoning engines.

b) **Biomedical Named Entity Recognition**
NER is a sub-task of natural language processing(NLP) that involves identifying and classifying named entities in a text into predefined categories such as person names,organizations,locations,medical codes,etc. It helps in extracting useful information from unstructured text data by identifying and classifying named entities and machine translation. Examples include: MED7, it is used for extracting the entities from the treatment charts.

c) **Distributional Hypothesis**
It refers to the concept where words that occur in similar contexts have similar meanings. Distributional representation generates a representation from a context while inducing distributional property. The surroundings with similar words may impart similar meaning to the word.

d) **HL7 Segments**
HL7 Segments are lines starting with a three-character identifier (Segment ID). Every line contains a group of fields containing various types of data. Each line is separated by a delimiter ("|"). All segments exist independently in the HL7 standard and can be utilized in the multiple messages, in various sequences, The first segment of any message is the MSH segment.

e) **Parse Trees in Biomedical NER**
Parse Treesrepresent a sentence in a directional way with nodes and branches. The tokens represent words and the edges represent syntactic relationships. It is hierarchical and uses rules based on grammar for parsing.There are two types of parsing: Constituency and Dependency Parsing. For example, sentences are broken down into entities which are categorized as noun or verb phrases and are further used to create a structure between these entities.

f) Stemming

Stemming is a natural language processing technique that is used to reduce the words to their base form, also known as the root form. The process of stemming is used to normalize the text and make it easier to process. There are several different algorithms for stemming: Porter Stemming, Snowball Stemming, Lancaster Stemming.

g) Lemmatization

Lemmatization is a process of text pre-processing that groups words according to its lemma(original word).It is done to identify the similarities among a group of words with the same root meaning.

SECTION-C [LONG ANSWER QUESTIONS] - 24 marks (8 marks x 3)

1. Create a storyboard for Eye Check Examinations for School Going Children in a Rural Indian Settings. [3 marks]. Create a UML diagram outlining the key actors and actions in the storyboard. [5 marks]

Answer: Storyboard should be clear with the problem statement and all the use-case details. UML diagrams should have all the major components defined with specific fields and variables used.

2. Enumerate at least three approaches to achieve Semantic Interoperability in Clinical text [3 marks]. Explain what is a Vector Space Embedding [5 marks]

Answer:

1.SNOMED-CT/ICD 2 . Biomedical Ontologies 3. Vector space embeddings

VECTOR SPACE EMBEDDINGS:

Vector space embeddings refers to the text that has been converted into numbers in a vector space. They capture the meaning and relationship between the words in the text. The dimensions of these vectors are independent of each other in a multidimensional space and help machine understand the text and process it more effectively. There are various types of vector embeddings such as word embedding, sentence embeddings and document embeddings etc.

Word embeddings represent individual words as vectors, for example, Word2Vec, GloVe, FastText.

Natural language processing uses vector embeddings for various tasks such as sentiment analysis, named entity recognition, clinical text classification, etc.

3. Explain how the FHIR framework enables to create an "internet of medical records" along with examples of resources. [3 marks]. Explain with an example profiling a FHIR resource? [5 marks]

Answer: RESTful APIs based framework. Resource implementation is the major concept that is used in RESTful APIs. Examples include:

a)Conformance: Resources describe how a system does or should work. Example: ValueSet, Conformance, StructureDefinition

b) Administration: Resources to manage the administrative side of healthcare. Example: Patient, Order, OrderResponse.

c) Clinical: Clinical summaries, record keeping and planning. Example: Observation, Condition, CarePlan, AllergyIntolerance.

d) Financial: Resources that support financial services associated with the provision of healthcare. Example: Claim, Coverage, ExplanationOfBenefit.

FHIR framework can be used in different categories to create an 'internet of medical records'. The structure of FHIR framework includes:

1. Conformance Framework
2. Infrastructure/API
3. Ontology
4. Workflow Management

Profiling of a FHIR resource includes:

- Value Sets: Provide a set of constraints on a resource for particular contexts of use.
- Specify a variety of restrictions such as - which of the values specified in the standards are allowed within a value set for the use case.
- For example: food, environment, biologics, medication, etc.