



Dưới đây là 20 câu hỏi trắc nghiệm về Xử lý Ngôn ngữ Tự nhiên

social psychonogy (Trường Đại Học Kỹ Thuật Công Nghiệp)



Scan to open on Studocu

1. **Tokenization là gì?**

- a) Phân chia văn bản thành các câu
- b) Phân chia văn bản thành các từ
- c) Loại bỏ stopwords khỏi văn bản
- d) Cả a và b**

2. **WordNet là gì?**

- a) Một từ điển thông thường
- b) Cơ sở dữ liệu từ vựng học cho tiếng Anh**
- c) Công cụ tìm kiếm
- d) Mô hình ngôn ngữ

3. **Synsets trong WordNet là gì?**

- a) Tập hợp các từ đồng nghĩa**
- b) Tập hợp các từ trái nghĩa
- c) Tập hợp các cụm từ liên quan
- d) Tập hợp các từ không liên quan

4. **Stemming là gì?**

- a) Tìm ngữ nguyên của từ
- b) Tìm từ đồng nghĩa
- c) Tìm từ trái nghĩa
- d) Loại bỏ đuôi của từ để đưa về gốc**

5. **Trong xử lý ngôn ngữ tự nhiên, collocations là gì?**

- a) Các từ xuất hiện gần nhau thường xuyên**
- b) Các từ không liên quan
- c) Các từ đồng nghĩa
- d) Các từ trái nghĩa

6. **Điểm tương đồng giữa hai synsets trong WordNet có thể được tính toán bằng phương pháp nào?**

- a) Jaccard Similarity
- b) Cosine Similarity
- c) Wu-Palmer Similarity**
- d) Dice Coefficient

7. **Trong NLTK, hàm nào được dùng để tokenize một đoạn văn bản thành các câu?**

- a) word_tokenize
- b) sent_tokenize**
- c) pos_tag
- d) ne_chunk

8. **Stopwords là gì?**

- a) Các từ quan trọng
- b) Các từ thường bị loại bỏ trong phân tích văn bản**
- c) Các từ hiếm gặp
- d) Các từ đồng nghĩa

9. **Trong phân tích văn bản, TF-IDF được dùng để làm gì?**

- a) Tìm từ đồng nghĩa
- b) Đo lường tầm quan trọng của từ trong một văn bản**
- c) Phân loại từ
- d) Tìm từ trái nghĩa

10. **POS tagging là gì?**

- a) Gán nhãn từ loại cho các từ trong câu**
- b) Phân chia câu thành các từ
- c) Tìm ngữ nghĩa của từ
- d) Tìm các từ đồng nghĩa

11. **Trong bài toán phân loại văn bản, bước đầu tiên thường là gì?**

- a) Tìm kiếm từ khóa
- b) Tokenization**
- c) Tìm ngữ nghĩa
- d) Gán nhãn từ loại

12. **Naive Bayes Classifier được sử dụng trong NLP để làm gì?**

- a) Phân loại văn bản**
- b) Tìm từ đồng nghĩa
- c) Tokenize văn bản
- d) Tìm từ trái nghĩa

13. **Sentiment analysis là gì?**

- a) Phân tích ngữ nghĩa
- b) Phân tích cảm xúc của văn bản**
- c) Phân loại từ
- d) Tokenize văn bản

14. **Trong phân tích sentiment, văn bản có thể được phân loại thành những nhóm nào?**

- a) Tích cực, tiêu cực, trung tính**
- b) Ngắn, dài, rất dài
- c) Dễ, khó, trung bình
- d) Đồng nghĩa, trái nghĩa, không liên quan

15. **Keyphrase extraction là gì?**

- a) Tìm kiếm các từ khóa quan trọng từ văn bản**
- b) Phân chia văn bản thành các từ
- c) Tìm từ đồng nghĩa
- d) Loại bỏ stopwords

16. **Feature extraction trong NLP là gì?**

- a) Loại bỏ các từ không quan trọng
- b) Tìm ngữ nghĩa của từ
- c) Trích xuất các đặc trưng từ văn bản**
- d) Tokenize văn bản

17. **Automated document summarization là gì?**

- a) Tóm tắt văn bản tự động**
- b) Phân chia văn bản thành các từ
- c) Loại bỏ stopwords
- d) Tìm từ đồng nghĩa

18. **Trong WordNet, hypernyms là gì?**

- a) Các từ cùng nghĩa
- b) Các từ trái nghĩa
- c) Các từ có nghĩa rộng hơn**
- d) Các từ có nghĩa hẹp hơn

19. **Công cụ nào trong NLTK giúp tìm kiếm các collocations?**

- a) BigramCollocationFinder**
- b) Synsets
- c) Lemmatization
- d) POS tagging

20. **Trong bài toán phân loại sentiment, đặc trưng nào có thể được sử dụng để huấn luyện mô hình?**

- a) Các từ thường xuyên xuất hiện**
- b) Các từ không xuất hiện
- c) Các từ đồng nghĩa
- d) Các từ trái nghĩa