

Анализ неструктурированных данных Самостоятельная работа Максимальная оценка: 10 баллов

Имя и фамилия: _____

Группа: _____

Ответьте на вопросы ниже. Можно использовать компьютер для вычислений и для просмотра лекций и учебников. Каждый вопрос оценивается в 2 балла.

1. Выберите верные утверждения:

1. В векторах word2vec / GloVe представлен один доминантный (самый частый) смысл слова;
2. Модель эмбедингов AdaGram помогает определить количество смыслов слова;
3. Если у слова два смысла (зАмок (Сущ.) и замОк (Сущ.)), половина компонент вектора word2vec отвечает за первый смысл, другая половина – за второй.
4. По одному вектору word2vec невозможно определить разные смыслы слова.

Ответ: 1, 2

2. Рассмотрим задачу определения смысла слова [word sense induction]. Методы кластеризации графов связей между словами в этой задаче используются:

1. Для определения кластеров синонимов (близких по смыслу слов) и антонимов (противоположных по смыслу слов);
2. Для определения кластеров гиперонимов (слов с более широким значением) и гипонимов (слов с более частным значением);
3. Для определения кластеров ближайших соседей;
4. Для определения однозначных и многозначных слов.

Ответ: 3, (4?)

3. Рассмотрим следующую постановку задачи WSI. Пусть дан корпус предложений. В каждом предложении встречаются многозначные слова и соответственно их контексты. Возникает следующая задача: по контекстам слова определить какой именно смысл слова использован в конкретном контексте. Назовите два этапа решения этой задачи и *очень* кратко опишите их.

Ответ: <http://panchenko.me/slides/graphs-lecture.pdf>, слайд 29

4. Что такое FrameNet?

1. Коллекция изображений в рамках и с подписями;
2. Нейронная сеть для sequence labelling (заполняет слоты конкретным смыслом слова);
3. База семантических фреймов (один фрейм: действие и его произвольное количество аргументов);
4. База синонимов и гиперонимов.

Ответ: 3

6. Как можно сделать интерпретируемыми эмбединги смыслов?

Ответ: <http://panchenko.me/slides/graphs-lecture.pdf>, слайд 54