

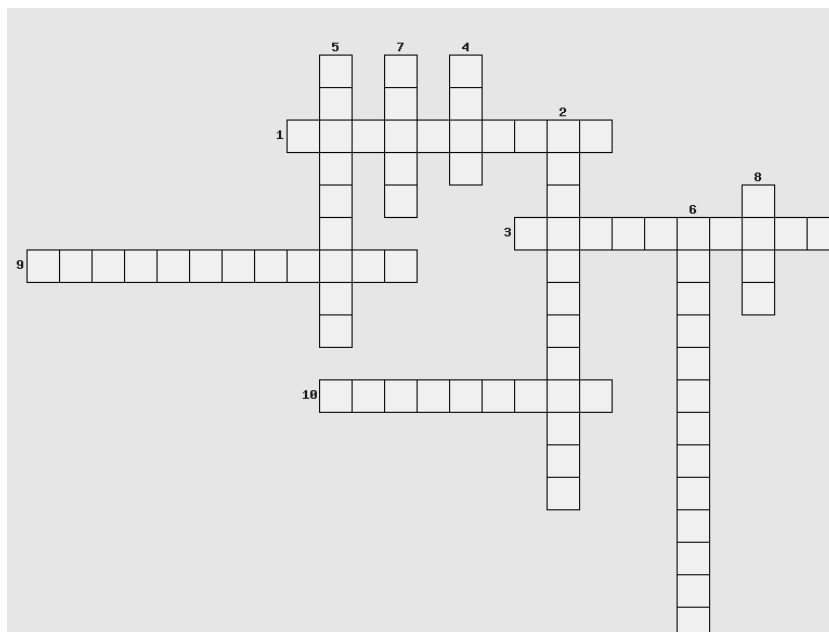
Вопросы и задания по докладу

Вопросы:

1. Что такое «граф»?
2. Что такое графовые методы в контексте анализа текста?
3. Какие задачи NLP можно решать с помощью графовых методов?
4. Как происходит предобработка текста перед построением графа?
5. Какие существуют основные этапы по созданию графов на основе текстовых данных?
6. Какие существуют метрики центральности? В чем их особенности?
7. В чем заключается задача «разрешения сущностей»?
8. Где применяются на практике «разрешения сущностей»?
9. В чем заключается визуализация и интерпретация графа?
10. Какие основные задачи визуализации графа?

Задания:

1. Решите кроссворд



По горизонтали:

1. Группа элементов, объединённых общими характеристиками или связями.
3. Вид графа, в котором одно и то же ребро может встречаться несколько раз.
9. Мера, отражающая степень разделения графа на подгруппы.
10. Обобщение графа, где рёбра могут соединять более двух вершин.

По вертикали:

2. Процесс представления информации в виде, удобном для зрительного восприятия.
4. Точка в графе, к которой присоединяются рёбра.

5. Характеристика, показывающая, насколько плотно заполнена структура графа рёбрами.
6. Объяснение или толкование смысла данных или результатов.
7. Элемент графа, который соединяет две вершины.
8. Структура данных, состоящая из множества вершин и соединяющих их линий.

2. Заполни таблицу:

Классификация графов в зависимости от типа взаимосвязей между узлами

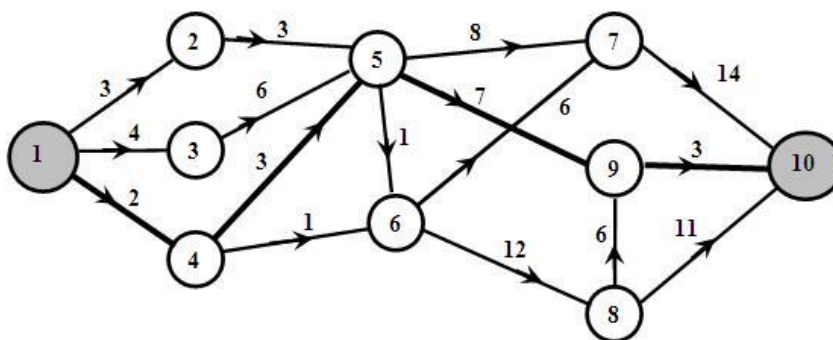
тип графа	тип взаимосвязи
ориентированные / направленные	
неориентированные / ненаправленные	
связные	
несвязные	
взвешенные	
невзвешенные	

3. Соотнеси метрики центральности с их объяснением:

1. центральность по степени
2. центральность по собственному вектору
3. центральность по близости
4. центральность по посредничеству

- А) учитывает «влиятельность» ближайших соседей, т.е. чем больше связей у соседних узлов, тем более важным считается данный узел
- Б) определяет количество кратчайших путей, проходящих через узел
- В) чем центральнее узел (т.е. чем короче путь от него до всех остальных узлов), тем выше его важность.
- Г) показывает важность узла с точки зрения количества связей с другими узлами в сети. Узел с большим числом связей считается более значимым.

4. Выясните како граф изображен на картинке. Для чего мог бы использоваться такой тип графа?



5. Разработайте алгоритм для анализа связности текста с использованием графовых методов.

Ответы на вопросы и решение заданий:

1. Что такое «граф»?

Граф — это математическая структура, представляющая собой множество вершин (узлов) и рёбер (связей), соединяющих эти вершины. Рёбра могут быть направленными (дуги) или ненаправленными.

2. Что такое графовые методы в контексте анализа текста?

Графовые методы в анализе текста — это подходы, использующие графы для моделирования и анализа текстовых данных. Они позволяют выявлять взаимосвязи между сущностями (например, словами, предложениями, документами) и анализировать их структуру.

3. Какие задачи NLP можно решать с помощью графовых методов?

- Определение семантической связанности текстов
- Анализ цитируемости и ссылок
- Выявление ключевых тем и сущностей
- Разрешение многозначности (disambiguation)
- Построение сетей цитирования и соавторства

4. Как происходит предобработка текста перед построением графа?

- Токенизация (разбиение текста на токены)
- Удаление стоп-слов
- Лемматизация и стемминг
- Синтаксический и морфологический анализ
- Векторизация (преобразование текста в числовой формат)

5. Какие существуют основные этапы по созданию графов на основе текстовых данных?

- Токенизация и выделение сущностей
- Создание матрицы смежности или списка рёбер
- Определение типов связей между сущностями
- Визуализация и анализ полученного графа

6. Какие существуют метрики центральности? В чем их особенности?

Степень центральности: количество связей узла

Межд�центральная центральность: кратчайший путь между узлами

Центральность по Кацу: учитывает все пути от узла

Центральность по собственному вектору: учитывает важность соседей

7. В чем заключается задача «разрешения сущностей»?

Задача разрешения сущностей заключается в идентификации и связывании упоминаний одних и тех же сущностей (например, людей, организаций, событий) в тексте.

8. Где применяются на практике «разрешения сущностей»?

- Обработка естественного языка и машинный перевод
- Рекомендательные системы
- Анализ социальных сетей
- Обработка юридических документов

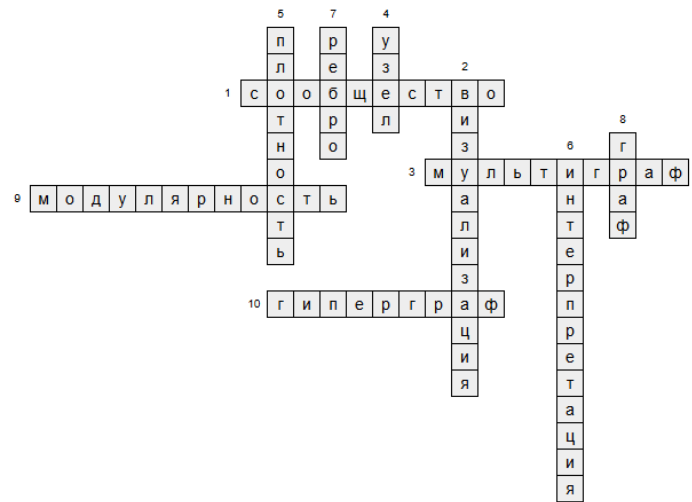
9. В чем заключается визуализация и интерпретация графа?

Визуализация графа — это процесс представления структуры графа в виде, удобном для восприятия. Интерпретация включает анализ выявленных связей и выявление закономерностей.

10. Какие основные задачи визуализации графа?

- Представление структуры данных наглядно
- Выявление ключевых узлов и связей
- Анализ центральности и влиятельности узлов
- Обнаружение сообществ и кластеров
- Визуализация динамики изменений в графах

1.



2.

тип графа	тип взаимосвязи
ориентированные / направленные	каждое ребро имеет направление, указывающее на то, какая вершина является начальной, а какая

	конечной
неориентированные / ненаправленные	ребра не имеют направления
связные	все узлы связаны между собой, что позволяет проложить как минимум один путь между любой парой вершин этого графа существует
несвязные	содержат узлы, которые оторваны от основного графа, то есть хотя бы две вершины не соединены
взвешенные	каждому ребру поставлено в соответствие некое числовое значение – вес ребра
невзвешенные	ребра не имеют веса

3.

-
- 1. центральность по степени
 - 2. центральность по собственному вектору
 - 3. центральность по близости
 - 4. центральность по посредничеству
- А) учитывает «влиятельность» ближайших соседей, т.е. чем больше связей у соседних узлов, тем более важным считается данный узел
 - Б) определяет количество кратчайших путей, проходящих через узел
 - В) чем центральнее узел (т.е. чем короче путь от него до всех остальных узлов), тем выше его важность.
 - Г) показывает важность узла с точки зрения количества связей с другими узлами в сети. Узел с большим числом связей считается более значимым.

4. На картинке изображен взвешенный ориентированный граф. Такие типы графов могут использоваться например:

- Для моделирования транспортных сетей, где рёбра могут представлять дороги или маршруты с весами, отражающими время в пути, расстояние или интенсивность движения.
- Для изучения взаимодействий между пользователями, где веса рёбер могут отражать частоту взаимодействий или уровень доверия между пользователями.
- Для моделирования зависимостей между задачами в проекте, где веса рёбер могут отражать время выполнения или ресурсы, необходимые для выполнения задач.

5. 4ок

1. Предварительная обработка текста

- Разбить текст на предложения .
- Привести к единому регистру, удалить знаки препинания, стоп-слова (при необходимости).
- Лемматизация / стемминг слов.

2. Векторизация предложений

Каждое предложение представляется в виде числового вектора. Можно использовать:

- TF-IDF
- Word2Vec / Перчатка
- BERT и другие трансформеры

3. Построение графа

Создать граф, где:

Вершины — предложения.

Ребра — это веса, зависящие от косинусного сходства между векторами предложений.

4. Оценка связности графа

1) Проверка графа на связность

2) Определение количества компонентов связности (чем больше компонентов, тем ниже связность.)

3) Центральность, PageRank, связность (эти показатели помогают выявить ключевые предложения, вокруг которых строится смысл)

5. Окончательная оценка связности на основе метрик

- Если граф связный и плотный, текст логичен и последователен.
- Если в тексте много связующих элементов, он может быть фрагментированным.
- Графовые метрики позволяют формализовать понятие «связности» текста и автоматизировать его оценку.