

Вариант 1.

ФИО Асирян А.К.

1. Перечислите и кратко охарактеризуйте виды моделей, используемых в КЛ. Для каждого вида укажите пример модели и прикладные задачи (1-2), в которых она используется.
2. Что такое **омонимия**? Какие виды омонимии Вы знаете? Как это явление влияет на автоматический анализ текста? Приведите примеры задач КЛ, в которых обязательно требуется проводить разрешение омонимии.
3. Для предложений укажите, являются ли они проективными, слабо проективными или непроективными. Ответ обоснуйте.
В лес пошли мы осенний.
Было очень красиво.
Мы много ягод собрали вкусных.
4. Что такое **морфологический анализ**? Что является выходом лемматизации? Приведите примеры задач КЛ, где лемматизация нужна, а где без нее можно обойтись. Ответ обоснуйте.
5. Что такое **анафора**? Как это явление влияет на автоматический анализ текста? Приведите примеры задач КЛ, в которых обязательно требуется проводить разрешение анафорических отсылок.
6. Зачем нужны корпуса и коллекции текстов? Что характерно для корпусов, какие их характеристики являются обязательными, а какие – факультативными?
7. Что такое **коллокация**? Приведите примеры. Укажите те характеристики коллокаций, которые помогают при их автоматическом выявлении.
8. Приведите примеры ситуаций, когда предложения без оценочных слов содержат оценку. Предложите метод обработки таких ситуаций.

насткое определение

звучание, но

характеристики являются обязательными, а какие — факультативными?

7. Что такое **коллокация**? Приведите примеры. Укажите те характеристики коллокаций, которые помогают при их автоматическом выявлении.

8. Приведите примеры ситуаций, когда предложения без оценочных слов содержат оценку. Предложите метод обработки таких ситуаций.

настоящее определение

2.) Омонимия — два слова сходны по написанию и звучанию, но различны по смыслу. Лексическая, грамматическая и синтаксическая (здесь речь о преф.)

Задачи:

• различия по ч.р. и по знач. морф. различия;

• семант. анализ.

~~грамм. анализ~~

• фонетический анализ

• машин. перевод

• По параметрам — различие морфемной грам. пар-в, но одинаковая лексика и ч.р.

тёплой утенок
им. п., м.р.

тёплой дорогой (дорога)
н.п., ж.р.

• По ч.р. — совп. по лексике, но разные ч.р. и пар-и
дорогой — прим. или сущ.

Дорогой, что такое омонимия?

• По лексике — совп. по ч.р., но разные по лексике.
красного вина
вино-сущ., р.п., ср.р.

мое вино
вино-сущ., им.п., ж.р.

• По ч.р. и лексике — ~~совп.~~ разные и по ч.р., и по лексике.
солнце пекло и попал в пекло
печь — м., ср.вр. пекло — сущ., ср.р.

4.) Морф. анализ - изучение лексик. и морф. параметров слова.
Лексикализация приводит слово к нач. форме. (форма ^{лекс.} → фр)
Куркина: подсчет вхождения слова, выделение ключ. слов + 4P
Можно обойтись: токенизация, сегментация ← в каком смысле
это преобразит. этап

5.) ~~Анафора - повторение слов, стоящих друг за другом в тексте~~
~~каждый инвертирует.~~

Анафора - интерпретация выражения зависит от другого, ранее встречавшегося в тексте.

Александр писал не то определение. Потом он исправил его.

Уточняется анализ текста. Например, в семант. анализе:

"кто исправил?", "что исправил?"

в выг. им. сущ.: "он" ~~скажем~~ не выводится как И.С. без
кон. анализа.

8.) "В отеле не было Wi-Fi." (гостиниц)

Можно составить словарь терминов для предг. анализа (РД), которые
передают хар-ки, например, для РД "отель": Wi-Fi,
телевизор, персонал, цена, еда. Также важно - сущ.

Потом необходимо определять оценку аспекта: мелкий Wi-Fi,

высшая еда. Для такого случая нужен словарь оценок слов
"не было Wi-Fi" нужно знать

передаются кар-ки, например, для ПО ~~оценки~~ "оценки": Wi-Fi, телевизор, персонал, цена, еда. Также важно - звук.

Потом необходимо определять оценочные аспекты: медленный Wi-Fi,

выручка еда. Для такого случая нужен словарь оценочных слов

для данной ПО. Для случая не было Wi-Fi" нужно знать

какие действия связаны с аспектами: Wi-Fi (есть/нет) -

- был, ловил, работал, пронад. Также нужно учитывать

комб.: "Wi-Fi медленно работал".

3.)

~~В лес пошел мы охотились - слабая проекция (покрытие корня)~~

~~Было очень плохо. - проективно~~

~~Мы много много собрали выручки. - не проективно (пересечение)~~

~~В лес пошел мы охотились. - не проективно (пересечение)~~

7.)

Комбинация - комбинация двух и более слов, имеющих тенденцию к совместной встречаемости. Белый дом, Владимир Путин, вынося приговор

частная встречаемость помечается и используется во всех ал-х.

ант. сп-ра

6.) ✓ • Покупка сырья - хромат

Киреев А.И.

• Создание сырьевых моделей

• Посредство договоров, договоров

• Обучение и тестирование морф. признаков и прикладных систем

~~Корпуса~~ Корпуса состав. решаемых задач, имеют большой объем (наборы), размещены по определенным правилам, естественно могут быть не связаны между собой.

Обяз: Размещение мн. решаемых задач
Представительская

Корпус: Сбалансированная
Масштабированная

1.) ✓ • Сырьевые модели

• учет качества изготовления единицы ← каких?
• учет посредственных единиц

Задачи: • мат. перевод
• морф. размещение

Н-графиковые модели чего?

• Структурная модель

• учет ~~структур~~ мн. особенностей одной/нескольких уровней ~~структур~~ и ~~связей~~

Задачи: • генерация текста
Набор мн. прав чего?

• Признаковая модель

• учет мн. и сырьевых ~~свойств~~ (признаков) сырья
↑
каких?

"Миним. слов"

Задачи: • классификация текстов
• классификация текстов

9. Для машинного перевода, основанного на правилах, используется словарь, где содержатся пары "слово-его перевод". Используется ли такого вида словарь для перевода, основанного на статистике? Если да, объясните для чего. Если нет, то каким образом тогда осуществляется перевод?

10. Перечислите основные приемы, на которых строится диалог в программах, моделирующих понимание собеседника, чат-ботах (от Элизы до Мишуки). Приведите примеры применения этих приемов.

10.) • Шаблоны: "привет, % имя %!"

• Использование диалогов:

Бот: как дела?

Ч: нормально

Ч: как дела?

Бот: нормально

• Тематическое понимание: Ч: сегодня был плохой день

Бот: грустно. Согласен

• Рефразирование: Ч: у меня болит голова

Бот: оу у тебя болит голова.
мобиль

• Сопоставление: Ч: грустно все это

Бот: и не говори

• Разумение и продолжение диалога: Ч: как дела?

Бот: нормально, а у тебя?

• Контрастные реплики: Ч: а мне понравилось

Бот: конечно

• Переадресация вопроса: Ч: чем занимаешься?

Бот: а ты?

9.)

Да, для лучшего понимания при выравнивании параллельных текстов может использоваться двуязычный словарь. Если расширивать сам алгоритм перевода, то нет, т.к. сам используется статистика (например, n-граммы).