

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. Что такое **морфологический анализ**? Что является выходом **стемминга**? Привести примеры задач КЛ, где стемминг нужен, а где без него можно обойтись. Ответ обоснуйте.

4. Для предложений укажите, являются ли они проективными, слабо проективными или непроективными. Ответ обоснуйте.

Мы в лес пошли осенний.

Красиво было очень.

Мы собрали много вкусных ягод.

5. Что такое **эллипсис**? Как это явление влияет на автоматический анализ текста? Приведите примеры задач КЛ, в которых обязательно требуется проводить разрешение эллипсиса.

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 оценочные слова не несут оценки. Предложите метод обработки таких ситуаций.

~~1) это частный случай~~ 2) нужно анализ омонимии (как я вписала)
Омонимия: слова, одинаково звучащие, но имеющие разный смысл.
Примеры: "коса", "наред"
Как это явл. влияет на ААТ?

Омонимия:

- лексическая (совпадение форм разных лексем)
- грамматическая (совпадение форм одной лексемы)
- синтаксическая (совпадение разных синтаксич. структур на основе одной послед-ти языковых знаков)

На этапе морфол. анализа текста ~~нужно~~ (сфер. морф. хар-к) необходимо учитывать омонимии в задачах (примеры):

- машинный перевод
- анализ тональности

Система синонимов - одна из самых сложных и важных частей морф. анализа
~~нужно~~ если ее необходимо проводить.

③ Морфологический анализ - этап обработки текста, при котором
определяются морфологические признаки (хар-ки) ~~лексемы~~ ^{лексем} (определение lemma, ~~часть~~ части речи и т.д.)

Стеминг - отсечение окончания слова по правилам

Вывод - основа/постфикс слова

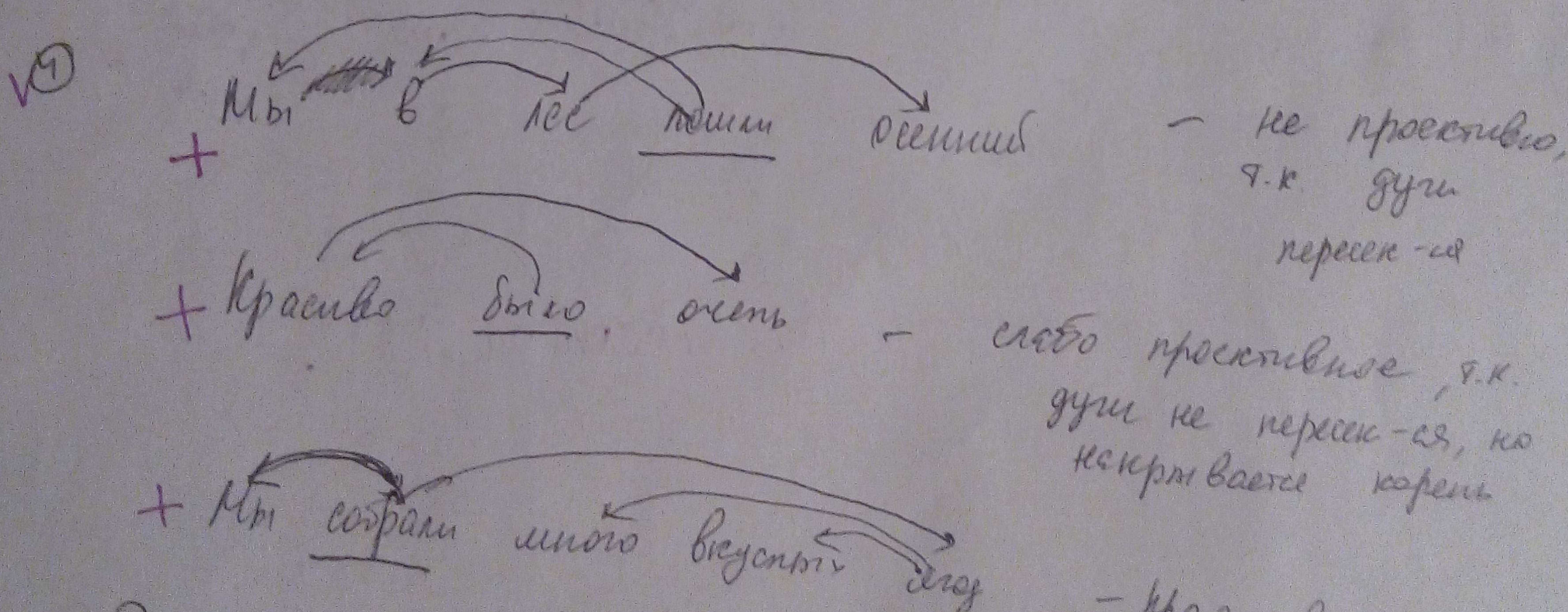
Примеры задач, где необходим стеминг:

- поиск информации (IR)
- ~~нужно~~ классификация/кластеризация текстов

Зависит от постановки задачи

В обеих задачах необходимо оценивать сходство смысла слов, но не важно, в какой форме слово встречается в тексте.

~~Без стеминга можно обойтись вообще~~ Без стеминга можно обойтись вообще
интересный пример, но в том случае морф. анализ вообще не нужен
задачах, где используется словарь речи (такие обратные процессы - из основы слова получаем словоформа.



⑤ Эллипсис - пропуск подразумеваемой единицы, который может быть восстановлен:
"Я поеду завтра, а он - на работу"
подразумевается "поедет"

Голосина Екатерина

⑤ Эмпатия вытекает из определения связности предложений, ~~предложения~~ ^{предложения не только в том, но и в}
Она может быть восполнена на основе слов, которые есть в
соседних предложениях ~~и в~~ (частях предложения)

Разрешение эмпирии можно, ~~например~~ ^{например}, при машинном
перевде (иначе смысл предложения теряется)

⑥ Корпуса и коллекции текстов представляют собой массивы
языковых данных. Используются в различных задачах,
~~например~~ особенно при работе со статистик. языковой
моделью: ~~например~~ ^(языковая модель)

- для получения языковых статист. данных
- при создании стат. моделей
- при обучении и тестировании программ

Примеры задач: оценка точности, чтение аннотаций,
маш. перевод.

Характеристики корпуса:

- разметка
- кол-во языков
- параллельность
- разделение по стилям/жанрам

это язык

Обязательная характеристика является наличие разметки
(в этом состоит основное ~~различие~~ ^{отличие} от коллекций)

! 1) создание для решения
опред. мн. задач

2) то, что и наличие корп. менеджера

3) естествен. контекста данных 4) представ. объем

это не слова, а их комбинации (словосочет.)
 ⑦ Коллокации - слова, которые часто используются друг с другом (устойчивое словосочетание)

Примеры: "вынести приговор", "сказать вкратце",
 "красное вино", "свежий ром"

Характеристики, кот. помогают при автоматич. выявлении:

- * - высокая частотность совместной встречаемости слов
- * - низкая частотность отдельной встречаемости слов
- * - название/имен - написание с заглавной буквы (не для всех языков)

это
фраза
(только для
устойчивых
фразеологизмов)

① Виды моделей ЕЯ и текста:

1) Статистические:

- опора на частоту употребл-я единиц и их посл-ств
- модель - единицы языка + частоты

~~пример модели:~~ какие?

Используются в таких задачах, как инф. поиск, определение авторства, оценка тематичности, кластер-я, класс-я

примеры моделей

2) Структурные:

- учитывает лингв. особенности одного или нескольких уровней ЕЯ или текста
- модель - набор лингв. правил. что?

Задачи: машинный перевод, генерация текста

3) Признаковые модели

- учитывают лингв. и статист. св-ва текста
- модель - признаки + веса

Задачи: извлечение инф-ции, проверка правописания, маш. перевод, генерация текста

не очень популярно, как там, и т.д.

Галкина Екатерина

Нет примеров

③

~

Оценочное слово может не вестись оценке в устойчивых словосочетаниях или термінах. Для решения

проблемы можно рассмотреть их ^{кого?} отдельно,

преждевременно их выделив из текста; или оценивать не отдельные слова, а словосочетания; или применить почему "им"?

различное обучение.

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 Мицуки). Приведите примеры применения этих приемов.

- ⑩
- Приемы
- 1) Использование ключевых слов
 - 2) Тактика активного слушания
 - 3) Неуровновешенная личность
 - 4) Шаблон
 - 5) Учет контекста
 - 6) Частичное понимание
 - 7) Нейтральные реакции
 - 8) Перефразирование
 - 9) Побуждение к продолжению диалога
 - 10) Отпаивание

Примеры:

1) You: - I had an argument with my mother ← ключевое слово из категории "семья"
Bot: - How much time do you spend with your family?

2) You: - There is something I want to talk about...
Bot: - Please go on. I want to know more.

3) Например, Bot Perry. На ~~и~~ имитацию неуравновешенной личности можно писать ошибочные фразы или выходящую за пределы тему.

You: - <some long phrase>

Bot: - I don't want to talk about it!

4) You: - I feel a bit bored.

Bot: - What makes you think you feel a bit bored?

выбирается одна из возможных фраз

5) Фразы "Earlier you said..." "ФРАЗА пользователя"

можно использовать по этим словам

6)

7) Bot: - This happens sometimes

Bot: - Oh, well.

8) Похоже на 4

You: - I like my job, but want more money

Bot: - So, you want to get more money for your work

9) Фразы: "Please go on", "Tell me more", "This seems interesting",
"Let's discuss it" и т.д.

10) Фразы "Ok", "You are right", "I agree"...

9) Да, такое слово может использоваться. ~~не~~ Возможные варианты использования:

- при выравнивании текста

- в диалоге, основанном на словах (может получиться несколько вариантов перевода; на основе статистических данных выясняется, как часто слова переводились именно так, если были в связи с группой слов или предложением; на основе

не совсем
понятно

подсчитанных значений выбирается предложение с лучшим переводом)