

Лекция 1

1. Что такое информационный поиск? Примеры поисковых систем
2. Связанные задачи с информационным поиском
3. Цель информационного поиска.
4. Модель поисковой системы
5. Задачи информационного поиска
6. Принципы булевого поиска
7. Как интерпретируется запрос в булевском представлении
8. Что такое ранжированный поиск
9. Как реализовать булев поиск
10. Пример матрицы «термин-документ»
11. Обратный индекс
12. Как реализуется обратный индекс
13. Как создается обратный индекс
14. Разбиение документа на токены
15. Словари и координатные блоки
16. Каковы затраты на хранение обратного индекса?
17. Какая логическая операция быстрее выполняется AND, OR, NOT?
18. Что такое прямое индекс?

Лекция 2

1. Чем занимается лингвистика
2. Зачем нужна фонетика в информационном поиске
3. Диахрония, синхрония
4. Чем занимается психолингвистика?
5. Порождающие грамматики.
6. Чем занимается количественная лингвистика, статическая лингвистика?
7. Рациональный и эмпирический подходы к языку.
8. Грамматика в терминах Хомского
9. Чем занимается морфология.
10. Аллморфы

Лекция 3

1. Что такое токен, термин?
2. Какие виды письменности существуют?
3. Как осуществляется детекция языка?
4. Принципы N-граммной детекции языка
5. Ранговый и марковский подходы анализа частотности языка
6. Пословный анализ частотности языка
7. Программные инструменты детекции языка
8. Нормализация терминов языка
9. Токенизация чисел

Лекция 4

1. Что такое коллокация, зачем их искать?
2. Признаки коллокации (некомпозиционность, незаменимость, немодифицируемость)
3. Частотность биграмм

4. Фильтр по частям речи в биграммах
5. Пример построения окна в биграммах
6. Рассчитать отклонение для устойчивого словосочетания
7. Построить гистограмму отклонений
8. Как проверяется гипотеза
9. Как осуществляется поиск коллокаций

Лекция 5

1. Что такое частота n-грам
2. Как осуществляется валидация моделей
3. Что такое кросс-валидаия
4. Линейное смешение моделей
5. Цепи Маркова, применимость в информационном поиске
6. Стохастическая матрица переходов
7. Графическое представление матрицы переходов
8. Скрытые марковские модели. Применение
9. Алгоритм "вперед" и алгоритм "назад"

Лекция 6

1. Что такое машинный перевод
2. Подходы к машинному переводу
3. Структура системы машинного перевода
4. Из чего состоит словарная статья
5. Уровни анализа предложения для перевода
6. Преимущества и недостатки машинного перевода на основе правил (RBMT)
7. Преимущества и недостатки статического машинного перевода
8. Что такое параллельный корпус
9. Источники параллельных корпусов
10. Выравнивание по предложениям
11. Функция выравнивания при машинном переводе
12. Ограничения в модели пословного перевода