

**федеральное государственное автономное образовательное учреждение высшего образования
Первый Московский государственный медицинский университет имени И.М. Сеченова
Министерства здравоохранения Российской Федерации
(Сеченовский Университет)**

Институт иностранных языков для профессиональных целей

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

Информационный поиск

основная профессиональная Высшее образование - бакалавриат - программа бакалавриата

45.00.00 Языкознание и литературоведение

45.03.04 Интеллектуальные системы в гуманитарной сфере

Профиль: Интеллектуальные системы в социологии и медицине

Интеллектуальные системы в лингвистике и медицине

**Раздел дисциплины (тема) 1: Теоретические основы информационного поиска.
Информационный поиск как область знания и практики**

Вид оценочного средства: устный опрос на знание терминологии

Уровень применения: Текущий контроль

№	Вопрос	Эталон ответа	Уровень применения
	Определите термин по его дефиниции.		
1.	Область знания и практики, изучающая и реализующая способы нахождения релевантной информации в ответ на информационную потребность пользователя	Информационный поиск	ТК
2.	Сфера теоретического осмысления явлений, понятий, закономерностей и методов, связанных с определённым предметом исследования.	Область знания	ТК
3.	Сфера применения методов, инструментов и технологий для решения реальных задач.	Область практики	ТК
4.	Исходные сведения, факты, сигналы или записи, ещё не интерпретированные в конкретном контексте.	Данные	ТК
5.	Осмысленные данные, которые приобрели значение для пользователя в определённой ситуации.	Информация	ТК
6.	Материальный или цифровой носитель зафиксированной информации, предназначенной для хранения, передачи и использования.	Документ	ТК
7.	Организованная совокупность документов, данных или сведений, доступная для поиска, использования и анализа.	Информационный ресурс	ТК
8.	Информационный ресурс, существующий в электронной форме и доступный через цифровые технологии.	Цифровой ресурс	ТК
9.	Совокупность документов, объединённых по теме, происхождению, типу, времени создания или другому признаку.	Документный массив	ТК
10.	Система, предназначенная для хранения, обработки, организации, поиска и предоставления информации	Информационная система	ТК

	пользователю.		
11.	Осознаваемая необходимость получить сведения для решения учебной, научной или практической задачи.	Информационная потребность	ТК
12.	Словесно или формально выраженная форма информационной потребности, вводимая пользователем в систему.	Поисковый запрос	ТК
13.	Конкретная цель поиска, например нахождение статьи, источника, цитаты, словарной статьи или примеров употребления.	Поисковая задача	ТК
14.	Процесс обмена действиями между пользователем и системой: формулирование запроса, получение выдачи, уточнение условий, оценка результатов.	Поисковое взаимодействие	ТК
15.	Средства взаимодействия пользователя с информационной системой: строка поиска, фильтры, рубрики, меню, результаты выдачи.	Поисковый интерфейс	ТК
16.	Список результатов, который система предлагает пользователю в ответ на запрос	Поисковая выдача	ТК
17.	Признаки, по которым пользователь или система ограничивает и уточняет поиск: тема, автор, дата, жанр, язык, тип документа.	Критерии отбора	ТК
18.	Перемещение пользователя по заранее организованной структуре ресурсов, разделов, коллекций или категорий.	Навигация	ТК
19.	Ограничение множества объектов по заданным параметрам без обязательного поиска нового содержания по смыслу.	Фильтрация	ТК
20.	Активное выявление информации, соответствующей запросу или потребности пользователя.	Поиск	ТК
21.	Поиск по словам и выражениям в полном тексте документа, а не только по его описанию или метаданным.	Полнотекстовый поиск	ТК
22.	Данные о данных, то есть сведения, описывающие документ или ресурс: автор, название, дата, ключевые слова, язык, источник.	Метаданные	ТК
23.	Специально организованное собрание цифровых объектов, объединённых общим принципом.	Цифровая коллекция	ТК

24.	Документ, текст, свидетельство или ресурс, содержащий сведения, используемые в исследовании.	Источник	ТК
25.	Электронное собрание текстов, организованное для поиска, анализа и изучения языка в употреблении.	Корпус	ТК
26.	Поиск информации по определённой теме, проблеме или предметной области.	Тематический поиск	ТК
27.	Продуманный способ организации поиска: выбор ресурса, ключевых слов, ограничений и способов уточнения запроса.	Поисковая стратегия	ТК
28.	Изменение исходного запроса для повышения точности и релевантности результатов.	Уточнение запроса	ТК
29.	Процесс поиска информации в информационных системах на основе сформулированного запроса.	Информационный поиск	ТК
30.	Формулировка пользователем информационной потребности в виде слов или выражений для поиска.	Поисковый запрос	ТК
31.	Набор слов или выражений, используемых для представления содержания документа.	Ключевые слова	ТК
32.	Процесс добавления документов в поисковую систему и создания их поисковых представлений.	Индексирование	ТК
33.	Совокупность документов, доступных для поиска в информационной системе.	Документная коллекция	ТК
34.	Поиск информации на основе сходства содержания документов и запроса.	Поиск по релевантности	ТК
35.	Результат работы информационно-поисковой системы, представляющий список найденных документов.	Поисковая выдача	ТК
36.	Упорядочивание найденных документов по степени их соответствия запросу пользователя.	Ранжирование документов	ТК
37.	Какой этап является первым в процессе информационного поиска?	Осознание информационной потребности	ТК
38.	Степень соответствия найденного документа информационной потребности пользователя.	Релевантность	ТК
39.	Показатель качества поиска, отражающий долю релевантных документов среди найденных.	Точность	ТК
40.	Показатель поиска, отражающий долю найденных	Полнота	ТК

	релевантных документов среди всех существующих релевантных документов.		
--	--	--	--

Раздел дисциплины (тема) 1: Теоретические основы информационного поиска.

Вид оценочного средства: тестовые задания

Уровень применения: Текущий контроль

№	Вопросы варианты ответа	Эталон ответа	Уровень применения
1.	Что в информационном поиске называется документом? А. Любой физический объект, содержащий текст В. Информационный объект, содержащий зафиксированную информацию, предназначенную для хранения и передачи С. Только печатная книга D. Только электронный файл	В	ТК
2.	2. Что означает термин <i>информация</i> в информационном поиске? А. Набор данных в компьютере В. Сообщение, уменьшающее неопределённость знаний пользователя С. Только научные публикации D. Любые тексты в интернете	В	ТК
3.	3. Что такое поисковый образ документа? А. Полный текст документа В. Формализованное представление содержания документа с помощью ключевых слов или индексов С. Список авторов документа D. Название файла	В	ТК
4.	4. Что называется поисковым образом запроса? А. Список ссылок на документы В. Формализованное представление информационной потребности пользователя С. Полный текст статьи D. База данных	В	ТК
5.	Что из перечисленного является показателем эффективности поиска? А. Скорость интернета В. Размер базы данных С. Точность и полнота поиска D. Количество серверов	С	ТК
6.	Что характеризует <i>точность поиска</i> ? А. Скорость выдачи результатов В. Долю релевантных документов среди найденных С. Общее количество документов D. Объём базы данных	В	ТК
7.	7. Что характеризует <i>полнота поиска</i> ? А. Долю найденных релевантных документов от общего числа релевантных документов в системе В. Количество найденных документов	А	ТК

	С. Скорость поиска D. Размер поискового индекса		
8.	Какой тип системы относится к специализированным научным базам данных? A. Google B. Scopus. C. Wikipedia D. YouTube	В	ТК
9.	Что является примером лингвистического информационного ресурса? A. Amazon B. British National Corpus C. Google Maps D. Instagram	В.	ТК
10.	10. Какой этап является первым в процессе информационного поиска? A. Анализ результатов поиска B. Формирование поискового запроса C. Осознание информационной потребности D. Сортировка документов	С	ТК
11.	9. Что является примером лингвистического информационного ресурса? A. Amazon B. British National Corpus C. Google Maps D. Instagram	В	ТК
12.	Какой тип системы является полнотекстовой поисковой системой? A. Scopus B. Web of Science C. Google D. электронный каталог библиотеки	С	ТК
13.	Что называется индексированием данных в информационном поиске? A. Процесс удаления данных из базы B. Процесс создания копий документов C. Процесс создания специальных структур данных, ускоряющих поиск информации D. Процесс сжатия файлов	С	ТК
14.	Основная цель индексирования данных — A. увеличение размера базы данных B. ускорение поиска и извлечения информации C. снижение качества поиска D. удаление дубликатов документов	В	ТК
15.	Что представляет собой поисковый индекс? A. Список пользователей системы B. Структура данных, связывающая термины с документами C. Каталог файлов D. База данных изображений	В	ТК
16.	Какой тип индекса используется для поиска по текстам? A. числовой индекс B. географический индекс C. полнотекстовый индекс	С	ТК

	D. табличный индекс		
17.	Что хранит полнотекстовый индекс? A. только названия документов B. слова и токены из текстового содержимого документов C. только метаданные D. список пользователей	B	TK
18.	Что происходит при полнотекстовом индексировании? A. тексты удаляются B. тексты разбиваются на слова (токены) C. тексты переводятся на другой язык D. тексты архивируются	B	TK
19.	Какое преимущество даёт индексирование? A. уменьшает количество документов B. ускоряет выполнение поисковых запросов C. увеличивает размер текста D. замедляет поиск	B	TK
20.	Какой недостаток индексирования упоминается чаще всего? A. уменьшение объёма данных B. увеличение требований к объёму хранилища C. невозможность поиска D. удаление документов	B	TK
21.	Почему индексирование может замедлять операции обновления базы данных? A. индекс уменьшает память B. индексы необходимо обновлять при добавлении или изменении данных C. индексы удаляют документы D. индексы уменьшают скорость интернета	B	TK
22.	Какой тип данных лучше всего подходит для полнотекстового индексирования? A. числовые данные B. графические данные C. текстовые данные D. видеофайлы	C	TK
23.	Почему полнотекстовое индексирование важно для корпусной лингвистики? A. оно уменьшает тексты B. оно позволяет быстро находить слова и контексты C. оно переводит тексты D. оно удаляет слова	B	TK
24.	Что происходит при поиске слова в корпусе языка? A. просматриваются все тексты B. поисковая система обращается к индексу C. тексты переводятся D. создаются новые документы	B	TK
25.	Что может быть барьером внедрения интеллектуального поиска? A. простота алгоритмов B. большой объём данных C. малое количество документов D. низкая скорость интернета	B	TK
26.	Почему многоязычные базы данных усложняют индексирование? A. они уменьшают объём данных	B	TK

	В. требуют обработки разных языков и морфологических систем С. не содержат текстов Д. не поддерживают поиск		
27.	Почему интеллектуальный поиск требует сложных алгоритмов? А. из-за маленького количества данных В. из-за необходимости анализа структуры и смысла текстов С. из-за отсутствия пользователей Д. из-за размера серверов	В	ТК
28.	Лингвист ищет исследования по теме метафоры в политическом дискурсе. Какой поисковый запрос будет наиболее точным? А. language politics В. metaphor politics С. political discourse metaphor Д. discourse	С	ТК
29.	Лингвист ищет статьи по теме корпусной лингвистики. Какой ресурс наиболее подходящий? А. YouTube В. Wikipedia С. GoogleScholar Д. Instagram	С	ТК
30.	Лингвист получил 100 результатов поиска. Из них 20 релевантны. Чему равна точность поиска? А. 10% В. 20% С. 50% Д. 80%	В	ТК
31.	В базе данных существует 50 релевантных документов. Система нашла 25 из них. Чему равна полнота поиска? А. 25% В. 50% С. 75% Д. 100%	В	ТК
32.	Лингвист хочет найти примеры употребления слова "sustainability" в академическом английском. Какой инструмент наиболее подходящий? А. Социальные сети В. Корпус языка С. Интернет-магазин Д. Форум	В	ТК
33.	Лингвист ищет статьи по теме semantic change. Он вводит запрос: semantic change OR semantic shift. Зачем используется оператор OR? А. Чтобы уменьшить результаты В. Чтобы исключить документы С. Чтобы найти документы с любым из терминов Д. Чтобы искать точную фразу	С	ТК
34.	Лингвист вводит запрос: "cognitivelinguistics". Что означает использование кавычек? А. Поиск отдельных слов В. Поиск точной фразы С. Поиск только в заголовке Д. Поиск только в библиографии	В	ТК
35.	Лингвист ищет статьи по теме pragmatics in medical communication.	В	ТК

	Первый запрос: communicationmedicine. Он получает много нерелевантных результатов. Как можно улучшить запрос? А. Удалить слова В. Использовать точные термины: "medicalcommunication"ANDpragmatics С. Использовать только одно слово D. Искать только по автору		
36.	У пациента наблюдаются следующие показатели: температура 39 °С, кашель, головная боль, потеря обоняния. Вопрос: Что здесь соответствует уровню "Информация"? А. 39 °С, кашель, головная боль В. Симптомы похожи на COVID-19 С. Назначить тест и самоизоляцию D. Снизить риск осложнений за счёт раннего вмешательства	В	ТК
37.	Ситуация: Студент регулярно опаздывает, не сдаёт задания вовремя, а в классе молчит. Что соответствует "знанию"? А. Он не отвечает на вопросы В. У него, возможно, трудности с адаптацией С. Дать ему письменную работу D. Перевести в другую группу	В	ТК
38.	Ситуация: Клиент снял 50 000 Р со счёта без объяснений. Что здесь является "данными"? А. Снятие 50 000 Р В. Клиент боится банкротства банка С. Консультант предлагает застраховать вклад D. Банк повышает лимит на снятие наличных	А	ТК
39.	Ситуация: Пост в Instagram набрал 15 000 лайков, 7000 репостов и 5000 комментариев. Что здесь — "информация"? А. 15 000 лайков В. Пост стал вирусным С. Использовать те же хэштеги D. Пост был опубликован в 20:00	В	ТК
40.	Ситуация: В реке зафиксированы высокие уровни аммиака и нитратов. Что здесь соответствует уровню "мудрости"? А. Превышены ПДК В. Вода потенциально опасна С. Надо ввести фильтрацию и ограничить сбросы D. Постоянно мониторить выбросы, вводить штрафы и обучать предприятия	Д	ТК

Раздел дисциплины (тема) 2: Модели информационного поиска, индексирование и языки запросов

Вид оценочного средства: тестовые задания

Уровень применения: Текущий контроль

№	Вопрос и варианты ответа	Эталон ответа	Уровень применения
1.	Поисковый образ запроса — это: А. Библиографическое описание статьи	В	ТК

	В. Формализованное представление информационной потребности пользователя С. Только заглавие документа Д. Архивная шифровка документа		
2.	Какое определение точнее всего передает суть формальной релевантности? А. Соответствие документа личным интересам пользователя В. Соответствие документа внешним признакам запроса С. Соответствие документа научной традиции Д. Соответствие документа только его названию	В	ТК
3.	Содержательная релевантность — это: А. Совпадение даты публикации с запросом В. Соответствие документа реальному смыслу и теме поисковой задачи С. Совпадение числа слов в запросе и документе Д. Наличие документа в каталоге	В	ТК
4.	Субъективная релевантность зависит прежде всего от: А. Цвета интерфейса системы В. Оценки полезности результата конкретным пользователем С. Длины текста Д. Формата файла	В	ТК
5.	Компромисс между полнотой и точностью означает, что: А. Эти показатели всегда одинаковы В. Повышение одного показателя нередко влияет на снижение другого С. Полнота важна только в библиотеке Д. Точность важна только в интернете	В	ТК
6.	Что обычно происходит при чрезмерном расширении запроса? А. Повышается только точность В. Уменьшается количество результатов С. Может повыситься полнота, но снизиться точность Д. Система перестает работать	С	ТК
7.	Что обычно происходит при сильном сужении запроса? А. Может повыситься точность, но снизиться полнота В. Полнота всегда повышается С. Релевантность исчезает Д. Все документы становятся одинаково полезными	А	ТК
8.	Какой этап поискового процесса является первым? А. Анализ результатов В. Формулирование задачи С. Уточнение стратегии Д. Сравнение найденных источников	В	ТК
9.	Что следует после формулирования задачи в типичном поисковом процессе? А. Выбор ресурса В. Публикация результатов	А	ТК

	С. Заккрытие поиска D. Архивация данных		
10.	Что предполагает анализ результатов? A. Только скачивание всех найденных файлов B. Оценку найденных материалов по степени их соответствия задаче C. Только копирование ссылок D. Удаление нерелевантных ресурсов из интернета	В	ТК
11.	Уточнение стратегии поиска необходимо для того, чтобы: A. Изменить тему исследования B. Улучшить результаты на основе анализа промежуточной выдачи C. Избежать чтения текстов D. Сократить размер библиотеки	В	ТК
12.	Что из перечисленного лучше всего иллюстрирует поисковую стратегию? A. Один случайно введенный запрос B. Общий план поиска с выбором ресурса, формулировкой запроса и его коррекцией C. Только чтение аннотации D. Только открытие первой ссылки	В	ТК
13.	Итеративный поиск — это: A. Поиск без изменений запроса B. Однократное обращение к одному документу C. Последовательный поиск с уточнением запросов и оценкой результатов D. Только поиск в алфавитном каталоге	С	ТК
14.	Что такое шум в поиске? A. Техническая ошибка клавиатуры B. Нерелевантные результаты, попавшие в выдачу C. Отсутствие интернета D. Повторное чтение одного текста	В	ТК
15.	Что такое потери в поиске? A. Удаленные пользователем закладки B. Документы, не сохраненные на компьютере C. Релевантные материалы, которые не были найдены системой D. Наличие лишних синонимов в запросе	С	ТК
16.	Какое утверждение о релевантности верно? A. Релевантность определяется только числом страниц B. Релевантность не связана с задачей пользователя C. Релевантность связана и с запросом, и с потребностью пользователя D. Релевантность одинакова для всех пользователей во всех случаях	С	ТК
17.	Какой пример лучше всего соответствует субъективной релевантности?	В	ТК

	А. Документ содержит слова из запроса В. Один и тот же результат полезен одному пользователю и бесполезен другому С. В документе указана дата Д. Система выдала результат первой строкой		
18.	Что означает формализация запроса? А. Перевод документа в PDF В. Представление потребности в форме, понятной поисковой системе С. Создание библиографической записи Д. Переписывание статьи вручную	В	ТК
19.	Какой показатель особенно важен, когда нужно найти как можно больше материалов по теме? А. Полнота В. Цветовая маркировка С. Объем кэша Д. Размер шрифта	А	ТК
20.	Какой показатель особенно важен, когда нужно быстро получить небольшое число наиболее точных результатов? А. Полнота В. Точность С. Формат файла Д. Скорость печати	В	ТК
21.	Что из перечисленного является примером уточнения запроса? А. Добавление тематического уточняющего слова В. Выключение компьютера С. Перенос файла в другую папку Д. Печать документа	А	ТК
22.	Что из перечисленного относится к этапам поискового процесса? А. Формулирование задачи, выбор ресурса, составление запроса, анализ результатов В. Только хранение документов С. Только чтение полного текста Д. Только просмотр изображений	А	ТК
23.	Какой вариант правильно описывает связь между запросом и потребностью? А. Они всегда полностью совпадают В. Запрос является формой выражения потребности, но может передавать ее не полностью С. Потребность всегда уже, чем запрос Д. Потребность не влияет на релевантность	В	ТК
24.	Что из перечисленного помогает повысить полноту поиска? А. Использование только одного очень узкого термина В. Добавление синонимов и близких по смыслу слов С. Удаление всех уточнений Д. Использование только точной цитаты без вариантов	В	ТК

25.	Что из перечисленного помогает повысить точность поиска? А. Добавление слишком общих слов В. Использование уточняющих терминов и ограничений С. Максимальное расширение запроса Д. Игнорирование результатов выдачи	В	ТК
26.	Релевантность результата окончательно оценивается: А. Только поисковым роботом В. Только алгоритмом ранжирования С. Пользователем в контексте своей задачи Д. Только библиотекарем	С	ТК
27.	Какое определение наиболее точно подходит для библиотечного каталога? А. Система поиска языковых примеров в корпусе В. Система описания и поиска библиотечных документов С. Интернет-энциклопедия Д. Система автоматического перевода	В	ТК
28.	Электронный каталог — это: А. Только список сайтов в интернете В. Цифровая форма библиотечного каталога С. Только база научных цитирований Д. Только корпус текстов	В	ТК
29.	Библиографическая база данных содержит прежде всего: А. Только изображения рукописей В. Структурированные сведения о публикациях С. Только полные тексты художественных произведений Д. Только устные выступления	В	ТК
30.	Полнотекстовая научная база отличается тем, что: А. Содержит только названия документов В. Дает доступ к полным текстам научных публикаций С. Работает только с карточками каталога Д. Включает только словарные статьи	В	ТК
31.	Веб-поисковая система предназначена прежде всего для: А. Поиска информации в интернете В. Хранения бумажных каталогов С. Разметки рукописей вручную Д. Издания словарей	А	ТК
32.	Корпусный ресурс особенно полезен для: А. Поиска реальных примеров употребления слов и конструкций В. Регистрации читательских билетов С. Покупки книг Д. Создания презентаций	А	ТК
33.	Словарный ресурс прежде всего используется для: А. Получения сведений о значении, употреблении и свойствах слов В. Подсчета цитируемости автора С. Замены корпуса	А	ТК

	D. Поиска только диссертаций		
34.	В чем основное различие между каталогом и полнотекстовой базой? A. Каталог всегда содержит аудиозаписи B. Полнотекстовая база дает доступ к тексту документа, а каталог часто только к его описанию C. Каталог работает только на бумаге D. Между ними нет различий	B	TK
35.	Какое утверждение о библиографической базе данных верно? A. Она обязательно содержит все полные тексты B. Она может содержать записи, аннотации и сведения о публикациях без полного текста C. Она предназначена только для школьников D. Она заменяет любую поисковую систему	B	TK
36.	Какой ресурс лучше выбрать для поиска статьи по автору, названию журнала, году издания? A. Библиографическую базу данных B. Только корпус C. Только словарь D. Только веб-поиск без ограничений	A	TK
37.	Какой ресурс лучше выбрать, если нужно найти реальные контексты употребления слова в публицистике? A. Корпусный ресурс B. Библиотечный каталог C. Карточный каталог D. Библиографический указатель без текста	A	TK
38.	Какой ресурс лучше выбрать, если нужно уточнить значение слова, его стилистическую помету и примеры? A. Словарный ресурс B. Только веб-поиск C. Только каталог D. Только база цитирования	A	TK
39.	Какой ресурс лучше выбрать, если нужно узнать, есть ли книга в фонде библиотеки? A. Библиотечный или электронный каталог B. Корпус C. Словарь D. Конкорданс	A	TK
40.	Что является ограничением веб-поисковой системы в учебно-научной работе? A. Избыточность и неоднородность результатов B. Невозможность найти сайты C. Отсутствие ключевых слов D. Невозможность вводить запрос	A	TK
41.	Что является преимуществом библиографической базы данных? A. Она всегда заменяет чтение источника	B	TK

	В. Она обеспечивает структурированный поиск публикаций С. Она нужна только библиотекарям Д. Она не требует оценки результатов		
42.	Что является преимуществом полнотекстовой научной базы? А. Возможность сразу работать с полным текстом публикации В. Возможность заменить все каталоги С. Наличие только кратких описаний Д. Отсутствие необходимости формулировать запрос	А	ТК
43.	Какое утверждение о корпусе верно? А. Корпус предназначен прежде всего для поиска книг по шифру В. Корпус — это структурированное собрание текстов для языкового анализа С. Корпус состоит только из словарных статей Д. Корпус не связан с поиском	В	ТК
44.	Конкорданс — это: А. Список библиотечных шифров В. Список контекстов употребления слова в корпусе С. Раздел словаря с грамматикой Д. Каталог журналов	В	ТК
45.	Какой ресурс менее всего подходит для поиска лингвистических примеров употребления слова? А. Библиотечный каталог В. Национальный корпус языка С. Корпусный ресурс Д. Электронный корпус текстов	А	ТК
46.	Что означает принцип выбора ресурса в зависимости от исследовательской задачи? А. Нужно всегда пользоваться только одной системой В. Следует выбирать такой ресурс, который лучше всего соответствует цели поиска С. Следует пользоваться только самыми популярными сайтами Д. Вид ресурса не влияет на результат	В	ТК
47.	Какой из перечисленных ресурсов является специализированным? А. Общая веб-поисковая система В. Корпус исторических текстов С. Обычная домашняя папка Д. Список закладок браузера	В	ТК
48.	Какое различие между веб-поисковой системой и научной базой данных является существенным? А. Научная база данных обычно более структурирована и ориентирована на академические публикации В. Веб-поиск не использует запросы С. Научная база не работает в интернете Д. Веб-поиск не выдает результатов	А	ТК
49.	Что из перечисленного лучше всего подходит для поиска научной литературы по теме исследования?	А	ТК

	А. Библиографическая база данных или полнотекстовая научная база В. Только словарь С. Только корпус Д. Только карточный каталог изображений		
50.	Какой ресурс особенно важен для поиска словарных данных? А. Электронный словарь В. Библиографическая база С. Веб-каталог магазинов Д. Каталог рукописных фондов	А	ТК
51.	Что из перечисленного относится к ограничениям корпуса? А. Он может быть ограничен по жанрам, времени и составу текстов В. Он всегда содержит все тексты языка С. Он всегда дает полные сведения об авторах Д. Он заменяет словари и каталоги	А	ТК
52.	Что из перечисленного относится к ограничениям каталога? А. Он может не предоставлять полный текст документа В. Он не позволяет искать вообще С. Он всегда нерелевантен Д. Он существует только офлайн	А	ТК
53.	Что из перечисленного лучше всего показывает различие между словарем и корпусом? А. Словарь дает нормативное и лексикографическое описание, а корпус — реальные контексты употребления В. Корпус содержит только заглавия С. Словарь не связан со словами Д. Между ними нет различий	А	ТК
54.	Для какой задачи веб-поисковая система может быть полезна в первую очередь? А. Для первичного ориентировочного поиска по широкой теме В. Только для поиска библиотечных шифров С. Только для лексикографического анализа Д. Только для поиска грамматических помет	А	ТК
55.	Какой ресурс наиболее уместен для поиска библиографического описания книги? А. Каталог или библиографическая база данных В. Только корпус С. Только словарь Д. Только конкорданс	А	ТК
56.	Какое утверждение верно? А. Все информационно-поисковые системы одинаковы по функциям В. Разные типы систем имеют разные возможности и ограничения С. Каталоги всегда лучше корпусов Д. Веб-поиск заменяет все специализированные ресурсы	В	ТК

57.	Что наиболее точно характеризует информационный поиск? А. Только хранение документов в электронном виде В. Область знания и практики, связанная с нахождением релевантной информации С. Исключительно технический процесс сканирования архивов Д. Только поиск по библиотечному каталогу	В	ТК
58.	Информационный поиск в системе цифровых гуманитарных дисциплин рассматривается как: А. Часть работы с цифровыми текстами, источниками и ресурсами В. Исключительно раздел программирования С. Только библиотечное дело Д. Только статистическая обработка чисел	А	ТК
59.	Что из перечисленного лучше всего соответствует понятию «данные»? А. Уже объяснённый вывод исследования В. Научная интерпретация источника С. Исходные сведения до их осмысления в контексте Д. Полностью оформленный документ	С	ТК
60.	Информация — это: А. Осмысленные данные, значимые для пользователя В. Любая папка на компьютере С. Только цифровое изображение Д. Набор символов без значения	А	ТК
61.	Документ — это: А. Только бумажная книга В. Только архивная рукопись С. Любой текст в памяти человека Д. Носитель зафиксированной информации	Д	ТК
62.	Что является информационным ресурсом? А. Только научная статья В. Организованная совокупность документов или данных, доступная для использования С. Только электронная таблица Д. Любой случайный набор файлов	В	ТК
63.	Цифровой ресурс — это: А. Любой устный рассказ В. Только аудиозапись С. Информационный ресурс, существующий в электронной форме Д. Только сайт университета	С	ТК
64.	Документный массив — это: А. Один документ большого объёма В. Совокупность документов, объединённых определённым признаком С. Только архив старопечатных книг Д. Любая пользовательская папка на ноутбуке	В	ТК
65.	Что отличает хранение информации от поиска?	А	ТК

	A. Хранение связано с сохранением, а поиск — с выявлением нужной информации B. Хранение всегда требует запроса, а поиск не требует C. Поиск используется только в архивах D. Между ними нет различия		
66.	Навигация — это: A. Автоматическое удаление нерелевантных документов B. Перемещение по структуре разделов, рубрик или коллекций C. Только поиск по ключевым словам D. Только сортировка по дате	B	TK
67.	Фильтрация — это: A. Перевод текста в цифровую форму B. Чтение всех найденных документов подряд C. Ограничение результатов по заданным параметрам D. Составление библиографии вручную	C	TK
68.	Собственно поиск — это прежде всего: A. Активное нахождение информации в ответ на потребность пользователя B. Создание цифровой копии документа C. Хранение файлов в облаке D. Простое перелистывание страниц	A	TK
69.	Поиск как взаимодействие пользователя с информационной системой означает, что: A. Пользователь только наблюдает за работой системы B. Система полностью заменяет исследователя C. Пользователь и система обмениваются действиями: запросом, результатом, уточнением D. Поиск происходит без участия человека	C	TK
70.	Что является первым шагом поискового процесса? A. Публикация найденных результатов B. Возникновение информационной потребности C. Составление списка литературы D. Сканирование документов	B	TK
71.	Как называется словесное выражение информационной потребности в поисковой системе? A. Метаданные B. Аннотация C. Фильтр D. Запрос	D	TK
72.	Что из перечисленного наиболее типично для информационного поиска гуманитария? A. Поиск только числовых таблиц B. Поиск только программного кода C. Поиск текстов, источников, примеров употребления и научной литературы D. Поиск только изображений без контекста	C	TK

73.	Поиск примеров употребления нужен гуманитариям прежде всего для того, чтобы: А. Проверить, как языковая единица функционирует в реальных контекстах В. Найти только дату публикации текста С. Определить объём компьютерной памяти D. Заменить чтение источников	А	ТК
74.	Словарные данные включают: А. Только фамилию автора словаря В. Значения слов, грамматические характеристики и примеры С. Только количество страниц D. Только место издания	В	ТК
75.	Корпус в гуманитарных исследованиях — это: А. Здание архива В. Тип библиотечного каталога С. Электронное собрание текстов для поиска и анализа D. Только собрание рукописей на бумаге	С	ТК
76.	Корпусный пример — это: А. Любой случайный фрагмент из интернета В. Фрагмент текста, найденный в корпусе и иллюстрирующий употребление единицы С. Ошибка в словаре D. Название корпуса	В	ТК
77.	Что из перечисленного лучше всего описывает релевантность? А. Объём документа в страницах В. Количество файлов в коллекции С. Соответствие результата запросу и информационной потребности D. Наличие цветной обложки	С	ТК
78.	Что из перечисленного относится к научной литературе? А. Монографии, статьи, диссертации В. Только личная переписка С. Только газетные объявления D. Только устные рассказы	А	ТК
79.	Какой пример лучше всего иллюстрирует навигацию, а не поиск? А. Ввод запроса в поисковую строку В. Переход по разделам «Литература → Языкознание → Корпусные исследования» С. Ввод слова в корпус и получение контекстов D. Поиск цитаты по фразе	В	ТК
80.	Какой пример лучше всего иллюстрирует фильтрацию, а не поиск? А. Выбор в базе данных только статей за 2020–2025 годы на английском языке В. Формулирование новой исследовательской проблемы С. Чтение найденной монографии	А	ТК

	D. Интерпретация текста источника		
81.	Что из перечисленного является цифровым ресурсом? A. Электронная библиотека статей B. Устная лекция без записи C. Разговор в коридоре D. Незаписанное интервью	A	TK
82.	Какое утверждение верно? A. Данные и информация — полностью тождественные понятия B. Документ всегда существует только на бумаге C. Цифровой ресурс может включать документы, данные и метаданные D. Поиск не связан с пользователем	C	TK
83.	Что является объектом поиска гуманитария чаще всего? A. Только технические характеристики оборудования B. Тексты, источники, словарные материалы, контексты употребления C. Только таблицы с измерениями D. Только математические формулы	B	TK
84.	Какова роль пользователя в информационном поиске? A. Только хранить найденные файлы B. Только запускать компьютер C. Формулировать потребность, задавать запрос и оценивать результаты D. Полностью отсутствовать в процессе	C	TK
85.	Что лучше всего показывает различие между поиском и навигацией? A. Поиск всегда медленнее навигации B. Навигация опирается на структуру ресурса, а поиск — на запрос пользователя C. Навигация используется только в печатных книгах D. Поиск невозможен в цифровой среде	B	TK
86.	Что из перечисленного не относится к специфике гуманитарного поиска? A. Поиск источников B. Поиск примеров употребления C. Поиск научной литературы D. Поиск только системных файлов компьютера	D	TK
87.	Метаданные — это: A. Только скрытый текст документа B. Сведения, описывающие документ или ресурс C. Исключительно цифровые изображения D. Только шифры архивных дел	B	TK
88.	Что наиболее точно описывает роль информационной системы в поиске? A. Она только хранит файлы и ничего больше не делает B. Она заменяет исследовательскую интерпретацию полностью	C	TK

	С. Она организует, обрабатывает и предоставляет результаты по запросу D. Она нужна только программистам		
89.	Что является лучшим примером документного массива ? A. Один словарь B. Совокупность статей журнала за несколько лет C. Один абзац текста D. Один поисковый запрос	В	ТК
90.	Почему для гуманитариев особенно важен поиск в корпусах и словарях? A. Потому что он позволяет работать с языком в употреблении и с зафиксированными значениями B. Потому что он полностью исключает необходимость анализа C. Потому что он заменяет все другие источники D. Потому что он нужен только программистам-лингвистам	А	ТК

Раздел дисциплины (тема) 2: Модели информационного поиска, индексирование и языки запросов

Уровень применения: Текущий контроль

Вид оценочного средства: Задачи с множественным выбором ответов

№	Вопрос и варианты ответа	Эталон ответа	Уровень применения
1.	Какой из примеров является примером поисковой задачи ? A. Найти корпусные примеры употребления слова в публицистике B. Установить операционную систему C. Создать резервную копию жёсткого диска D. Изменить яркость экрана	А	ТК
2.	Какой вариант верно отражает логику поискового взаимодействия? A. Выдача → документ → пользователь → запрос B. Запрос → информационная потребность → система → удаление данных C. Информационная потребность → запрос → выдача → оценка результатов D. Сканирование → хранение → печать → индексирование	С	ТК
3.	Расположите запросы в порядке возрастания количества страниц , которые найдет поисковый сервер. А) физкультура Б) физкультура & подтягивания & отжимания В) физкультура & подтягивания Г) физкультура фитнес Варианты ответа: А) Б → В → А → Г	А	ТК

	В) $B \rightarrow B \rightarrow A \rightarrow \Gamma$ С) $B \rightarrow B \rightarrow \Gamma \rightarrow A$ D) $B \rightarrow B \rightarrow \Gamma \rightarrow A$		
4.	Расположите запросы в порядке возрастания количества найденных страниц . А) музыка классика Моцарт серенада Б) музыка классика В) музыка классика Моцарт Г) музыка & классика & Моцарт Варианты ответа: А) $\Gamma \rightarrow Б \rightarrow В \rightarrow А$ В) $Б \rightarrow \Gamma \rightarrow В \rightarrow А$ С) $\Gamma \rightarrow В \rightarrow Б \rightarrow А$ D) $В \rightarrow Б \rightarrow \Gamma \rightarrow А$	А	ТК
5.	Расположите запросы в порядке возрастания количества результатов поиска . А) реферат математика Гаусс Б) реферат математика Гаусс метод В) реферат математика Г) реферат & математика & Гаусс Варианты ответа: А) $\Gamma \rightarrow В \rightarrow А \rightarrow Б$ В) $В \rightarrow \Gamma \rightarrow А \rightarrow Б$ С) $\Gamma \rightarrow А \rightarrow В \rightarrow Б$ D) $А \rightarrow В \rightarrow \Gamma \rightarrow Б$	А	ТК
6.	А) Гренландия & Климат & Флора & Фауна Б) Гренландия & Флора В) (Гренландия & Флора) Фауна Г) Гренландия & Флора & Фауна Варианты ответа: А) $А \rightarrow \Gamma \rightarrow Б \rightarrow В$ В) $Б \rightarrow \Gamma \rightarrow А \rightarrow В$ С) $А \rightarrow Б \rightarrow \Gamma \rightarrow В$ D) $\Gamma \rightarrow А \rightarrow Б \rightarrow В$	А	ТК
7.	Расположите запросы в порядке возрастания количества результатов поиска . А) (огурцы & помидоры) & (прополка поливка) Б) огурцы помидоры В) огурцы Г) огурцы & помидоры Варианты ответа: А) $А \rightarrow \Gamma \rightarrow В \rightarrow Б$ В) $\Gamma \rightarrow А \rightarrow В \rightarrow Б$ С) $А \rightarrow В \rightarrow \Gamma \rightarrow Б$ D) $\Gamma \rightarrow В \rightarrow А \rightarrow Б$	А	ТК
8.	Расположите запросы в порядке возрастания количества	А	ТК

	результатов. А) экзамен тестирование Б) (физика химия) & (экзамен тестирование) В) физика & химия & экзамен & тестирование Г) физика химия экзамен тестирование Варианты ответа: А) $B \rightarrow B \rightarrow A \rightarrow \Gamma$ В) $B \rightarrow B \rightarrow A \rightarrow \Gamma$ С) $B \rightarrow A \rightarrow B \rightarrow \Gamma$ Д) $B \rightarrow B \rightarrow \Gamma \rightarrow A$		
9.	Расположите запросы в порядке возрастания количества найденных страниц. А) сомики меченосцы содержание Б) сомики & содержание В) сомики & меченосцы & разведение & содержание Г) (сомики меченосцы) & содержание Варианты ответа: А) $B \rightarrow B \rightarrow \Gamma \rightarrow A$ В) $B \rightarrow B \rightarrow A \rightarrow \Gamma$ С) $B \rightarrow \Gamma \rightarrow B \rightarrow A$ Д) $\Gamma \rightarrow B \rightarrow B \rightarrow A$	А	ТК
10.	Даны документы: D1: «Одно из величайших заблуждений — это думать, что все чувствуют, видят и мыслят точно так же, как и мы.» D2: «Только тот, кто не страдал бессонницей, может думать, что все обстоит как нельзя лучше.» D3: «Спасать. Опять спасать. До каких же пор вас нужно будет спасать? Вы когда-нибудь научитесь спасать сами себя?» D4: «Цитаты — удобное средство не думать самому.» D5: «Люди не меньше привязываются к тем, кому сделали добро сами, чем к тем, кто сделал добро им.» D6: «Нет большей муки, чем воспоминание в несчастье о счастливом времени.» Какие документы удовлетворяют запросу: "не"AND"что" Варианты ответа: А) D1, D2 В) D2 С) D2, D4 Д) D1	В	ТК
11.	Даны документы: D1: «Одно из величайших заблуждений — это думать, что все чувствуют, видят и мыслят точно так же, как и мы.» D2: «Только тот, кто не страдал бессонницей, может думать, что все обстоит как нельзя лучше.» D3: «Спасать. Опять спасать. До каких же пор вас нужно будет спасать? Вы когда-нибудь научитесь спасать сами себя?»	А	ТК

	D4: «Цитаты — удобное средство не думать самому.» D5: «Люди не меньше привязываются к тем, кому сделали добро сами, чем к тем, кто сделал добро им.» D6: «Нет большей муки, чем воспоминание в несчастье о счастливом времени.» Какие документы удовлетворяют запросу: NOT ("не"OR"кто") Варианты ответа: A) D3 B) D1 и D3 C) D3 и D6 D) D4 и D6		
12.	Даны документы: Doc1 — breakthrough drug for schizophrenia Doc2 — new schizophrenia drug Doc3 — new approach for treatment of schizophrenia Doc4 — new hopes for schizophrenia patients Какой булев вектор соответствует слову new? Варианты ответа: A) 0111 B) 1011 C) 1101 D) 10019	A	ТК
13.	Какой запрос даст наибольшее количество результатов? A) language & AI B) language AI C) language & AI & linguistics D) language	B	ТК
14.	Какой запрос даст наименьшее количество результатов? A) data B) data & mining C) data & mining & algorithm D) data mining	C	ТК
15.	Какой запрос вернет наибольшее количество документов? A) медицина & диагностика B) медицина диагностика C) медицина & диагностика & лечение D) диагностика	B	ТК
16.	Даны документы D1: машинное обучение и анализ данных D2: методы анализа данных D3: глубокое обучение D4: алгоритмы машинного обучения Запрос обучение AND данные Какие документы будут найдены?	A	ТК

	A) D1 B) D1 и D2 C) D1 и D4 D) D2 и D3		
17.	Даны документы: Doc1: data mining techniques Doc2: data analysis Doc3: mining algorithms Сколько уникальных терминов содержит корпус? A) 4 B) 5 C) 6 D) 7	В	ТК
18.	Какой фактор НЕ используется в ранжировании поисковых систем? A) частота термина B) обратная частота документа C) длина документа D) цвет текста	D	ТК

Раздел дисциплины (тема) 3: Информационный поиск в гуманитарных и лингвистических исследованиях

Вид оценочного средства: Проектные задания

Уровень применения: Текущий контроль

№	Проектные задания	Уровень применения
1.	Поиск научной литературы по заданной теме Цель: Выполнить поиск научной литературы по предложенной гуманитарной или лингвистической теме с использованием электронных библиотек, каталогов, библиографических и полнотекстовых баз данных. Результат: подборка релевантных источников с указанием поисковых запросов и использованных ресурсов.	ТК
2.	Поиск источников по различным параметрам Цель: Осуществить поиск публикаций по автору, названию, теме, ключевым словам, году издания и типу публикации; сопоставить полученные результаты. Результат: краткий аналитический отчет о различиях в результатах поиска в зависимости от выбранного параметра.	ТК
3.	Сравнение библиографического и полнотекстового поиска Цель: Провести поиск по одной теме с использованием библиографических записей и полнотекстового поиска; определить различия в составе, объеме и характере результатов. Результат: сравнительная характеристика двух типов поиска с выводом об	ТК

	их учебной и исследовательской ценности.	
4.	Работа со словарными и лексикографическими ресурсами Цель: Осуществить поиск информации о заданной единице в различных словарных ресурсах; сопоставить толкования, примеры употребления, пометы и справочную информацию. Результат: краткое описание различий между словарными источниками и вывод о целесообразности их использования для решения разных исследовательских задач.	ТК
5.	Оценка качества и надежности источников Цель: Провести экспертную оценку найденных материалов по критериям авторитетности, научной надежности, актуальности, достоверности, проверяемости и происхождения данных. Результат: аналитическая карточка оценки источников с аргументированным выводом об их пригодности для учебной или научной работы.	ТК
6.	Систематизация найденной информации Цель: Сгруппировать найденные материалы по тематике, типу источника, степени научной значимости или исследовательской функции. Результат: структурированная тематическая подборка источников.	ТК
7.	Подготовка аннотированной библиографии Цель: Составить аннотированный список литературы по исследовательской теме на основе найденных и отобранных источников. Результат: библиография с краткими аннотациями, отражающими содержание и практическую ценность каждого источника.	ТК
8.	Оформление библиографического списка Цель: Подготовить библиографический список по заданному стандарту оформления на основе найденных источников. Результат: корректно оформленный список литературы для реферата, курсовой работы или исследовательского проекта.	ТК

Критерии оценивания

Оценка	Критерии оценивания
«Удовлетворительно»/ зачтено	Ответ полностью соответствует установленным критериям: демонстрирует глубокое понимание материала, логичность изложения, корректное использование терминологии, опору на изученные источники и способность к аргументированной самостоятельной интерпретации. Ответ в целом удовлетворяет предъявляемым требованиям, однако содержит отдельные недостатки: незначительные погрешности в отборе или анализе материала, отдельные нарушения логики изложения. Вместе с тем, общая логика и аргументация сохраняются.
Оценка «Неудовлетворительно» / не зачтено	Содержание источника понято в основных чертах, однако ответ строится преимущественно на репродуктивном воспроизведении материала с опорой на один источник. Самостоятельный анализ и обобщение отсутствуют.

	Ответ свидетельствует о неспособности студента к самостоятельному анализу материала: терминологический аппарат не используется или применяется некорректно, изложение отличается нелогичностью и неполнотой. Оценка выставляется также в случаях, когда ответ не соответствует поставленному вопросу либо обнаруживает полное непонимание темы.
--	--

Раздел дисциплины (тема) 4: Использование цифровых, программных и интеллектуальных средств в информационном поиске.

Уровень применения: Текущий контроль

Вид оценочного средства: Тестовые задания

№	Вопрос и варианты ответа	Эталон ответа	Уровень применения
1.	Что означает представление текста в виде данных? А. Перевод текста на другой язык В. Преобразование текста в структуру признаков, пригодных для обработки системой С. Сокращение текста до аннотации Д. Только сохранение текста в PDF	В	ТК
2.	Что такое терм в поисковом контексте? А. Любая строка интерфейса В. Единица поискового представления текста С. Только заголовок документа Д. Только научный термин узкой области	В	ТК
3.	Частота термина — это: А. Количество документов в библиотеке В. Число вхождений термина в документе или коллекции С. Число страниц в журнале Д. Позиция слова в словаре	В	ТК
4.	Вхождение — это: А. Любой найденный документ В. Отдельный случай появления единицы в тексте С. Название раздела каталога Д. Список авторов публикации	В	ТК
5.	Что в прикладном поисковом контексте называется словоформой? А. Только словарная статья В. Конкретная грамматическая форма слова С. Любое предложение Д. Тема текста	В	ТК
6.	Лемма — это: А. Первое слово в тексте В. Самая длинная словоформа С. Базовая словарная форма слова	С	ТК

	D. Любой знак препинания		
7.	Что такое токен? A. Единица текста, выделяемая системой при обработке B. Название базы данных C. Раздел тезауруса D. Только цифра в документе	A	ТК
8.	Токенизация — это: A. Перевод текста в изображение B. Разбиение текста на минимальные единицы обработки C. Удаление всех глаголов D. Сжатие документа	B	ТК
9.	Документ в поисковой системе рассматривается прежде всего как: A. Совокупность индексируемых признаков B. Только визуальный объект C. Последовательность страниц без признаков D. Только авторский текст без описания	A	ТК
10.	Что отражают частотные характеристики текста? A. Только жанр текста B. Количественное распределение единиц в тексте C. Только объем документа в байтах D. Только число абзацев	B	ТК
11.	Что лучше всего описывает связь между содержанием документа и его поисковым описанием? A. Поисковое описание всегда полностью тождественно содержанию B. Поисковое описание формально отражает некоторые существенные признаки содержания C. Содержания документа в поиске не существует D. Поисковое описание зависит только от цвета интерфейса	B	ТК
12.	Почему одна и та же информация может по-разному представляться в поисковом контексте? A. Только из-за технических ошибок B. Из-за лексической, морфологической и иной вариативности C. Только из-за длины документа D. Потому что система случайно меняет слова	B	ТК
13.	Модель «мешка слов» предполагает, что: A. Порядок слов строго важнее их состава B. Документ представляется как набор слов без учета их порядка C. Учитываются только заглавные буквы D. Учитываются только предложения, а не слова	B	ТК
14.	Что из перечисленного является причиной морфологической вариативности? A. Использование разных жанров B. Наличие у слова разных грамматических форм C. Наличие разных авторов D. Различие шрифтов	B	ТК

15.	Что из перечисленного ближе всего к понятию «словарь коллекции»? А. Список авторов коллекции В. Совокупность всех уникальных термов коллекции С. Каталог библиотек Д. Список только редких слов	В	ТК
16.	Какой вариант правильно различает слово и токен? А. Это всегда одно и то же В. Слово — языковая единица, а токен — единица, выделяемая системой при обработке С. Токен существует только в бумажных документах Д. Слово используется только в корпусах, а токен только в библиотеках	В	ТК
17.	Что из перечисленного помогает уменьшить вариативность представления слова в поиске? А. Нормализация и лемматизация В. Увеличение шрифта С. Изменение цвета текста Д. Добавление иллюстраций	А	ТК
18.	Что такое индекс в информационно-поисковой системе? А. Только классификационный шифр книги В. Структура данных для быстрого поиска документов С. Любая аннотация к статье Д. Перечень пользователей системы	В	ТК
19.	Индексирование — это: А. Процесс создания поискового представления документа В. Только чтение документа человеком С. Только перевод текста в другой формат Д. Только сохранение файла в базе	А	ТК
20.	Чем ручное индексирование отличается от автоматического? А. Ручное выполняется человеком, автоматическое — системой В. Ручное используется только для книг, автоматическое — только для статей С. Между ними нет различий Д. Автоматическое всегда точнее по смыслу	А	ТК
21.	Что такое дескриптор? А. Случайно выбранное слово текста В. Стандартизированный термин предметного описания С. Название поисковой системы Д. Формат файла документа	В	ТК
22.	Предметная рубрика — это: А. Нормативная тематическая формулировка для описания документа В. Только заголовок статьи С. Любой тег, добавленный пользователем Д. Раздел личной папки	А	ТК

23.	Что такое тезаурус в информационном поиске? А. Случайный список слов документа В. Система терминов и смысловых связей между ними С. Только частотный словарь Д. Только электронный каталог	В	ТК
24.	Что из перечисленного относится к метаданным? А. Автор, заглавие, дата, источник В. Только полный текст статьи С. Только частоты слов Д. Только логические операторы	А	ТК
25.	Поиск по метаданным — это: А. Поиск только по содержанию полного текста В. Поиск по описательным элементам документа С. Поиск по цвету обложки Д. Поиск только по иллюстрациям	В	ТК
26.	Что относится к преимуществам ручного индексирования? А. Более глубокий учет смысла документа В. Полное отсутствие трудозатрат С. Неограниченная скорость обработки Д. Полная объективность во всех случаях	А	ТК
27.	Что относится к преимуществам автоматического индексирования? А. Абсолютное понимание смысла текста В. Скорость и масштабируемость С. Полное отсутствие ошибок Д. Невозможность работать с большими массивами	В	ТК
28.	Что является ограничением автоматического индексирования? А. Высокая скорость В. Возможная потеря смысловых нюансов С. Возможность обработки больших коллекций Д. Использование алгоритмов	В	ТК
29.	Контролируемый словарь нужен прежде всего для: А. Полного исключения терминов В. Единообразного индексирования документов С. Создания случайных запросов Д. Увеличения длины аннотаций	В	ТК
30.	Что означает полнотекстовое индексирование? А. Индексирование только заглавий В. Индексирование слов и выражений из полного текста С. Индексирование только библиографических данных Д. Индексирование только предметных рубрик	В	ТК
31.	Что означает глубина индексирования? А. Размер экрана пользователя В. Степень подробности описания документа в индексе С. Высоту расположения записи в каталоге Д. Скорость интернет-соединения	В	ТК

32.	Какой вариант лучше всего описывает роль индекса? А. Индекс заменяет документ В. Индекс ускоряет поиск по признакам документа С. Индекс нужен только библиотекарям D. Индекс используется только вручную	В	ТК
33.	Какой из перечисленных элементов может быть идентификатором документа? А. DOI В. Часть абзаца С. Любое служебное слово D. Только номер страницы	А	ТК
34.	Что из перечисленного относится к ограничениям ручного индексирования? А. Скорость и масштабируемость В. Субъективность и трудоемкость С. Возможность учитывать контекст D. Использование профессионального опыта индексатора	В	ТК
35.	Что такое модель информационного поиска? А. Только интерфейс поисковой строки В. Способ представления документов, запросов и их сопоставления С. Только схема библиотеки D. Формат библиографического описания	В	ТК
36.	Что является слабой стороной булевой модели? А. Отсутствие точных условий В. Отсутствие градации сходства между запросом и документом С. Высокая прозрачность D. Использование операторов	В	ТК
37.	Векторная модель предполагает, что: А. Документ не содержит термов В. Документ и запрос представляются как наборы признаков с весами С. Поиск всегда неранжированный D. Используются только предметные рубрики	В	ТК
38.	Что в векторной модели понимается под сходством документа и запроса? А. Совпадение только длины текста В. Степень близости их признаковых представлений С. Совпадение формата файла D. Совпадение даты создания	В	ТК
39.	Почему векторная модель удобна для объяснения ранжирования? А. Потому что она вообще не сравнивает документы В. Потому что документы могут быть ближе или дальше от запроса С. Потому что она исключает запрос пользователя D. Потому что в ней нет весов термов	В	ТК

40.	Вероятностная модель исходит из идеи, что система должна: А. Искать только точные фразы В. Оценивать вероятность релевантности документа С. Полностью игнорировать статистику D. Искать только по заголовкам	В	ТК
41.	Что означает «предугадывать полезность» в вероятностной модели? А. Случайно выбирать документы В. Оценивать, какие документы с большей вероятностью будут полезны пользователю С. Показывать только новые документы D. Исключать все длинные тексты	В	ТК
42.	Какой тип выдачи наиболее характерен для векторной модели? А. Ранжированный список по степени сходства В. Алфавитный список авторов С. Нумерация документов по дате D. Только бинарное «да/нет»	А	ТК
43.	Какой тип выдачи наиболее характерен для булевой модели? А. Набор документов, соответствующих точным логическим условиям В. Обязательно вероятностное ранжирование С. Список словоформ D. Только фразовые совпадения	А	ТК
44.	Какая модель особенно полезна для учебного и библиотечного поиска при четко заданных условиях? А. Булева В. Только вероятностная С. Только корпусная D. Только векторная без операторов	А	ТК
45.	Что означает модель частичного совпадения? А. Документ может соответствовать запросу в разной степени В. Запрос всегда ошибочен С. Документ должен совпадать с запросом полностью D. Система не использует признаки	А	ТК
46.	От чего зависит характер поисковой выдачи? А. Только от скорости интернета В. В том числе от модели поиска, лежащей в основе системы С. Только от длины документа D. Только от размера шрифта	В	ТК
47.	Какой запрос лучше показывает потребность в векторной или вероятностной модели, а не только в булевой? А. Найти документы, где есть ровно одна заданная фраза В. Найти наиболее близкие по теме тексты о языковой вариативности С. Найти запись по точному DOI D. Найти книгу по инвентарному номеру	В	ТК

48.	Что является общим для векторной и вероятностной моделей? А. Полное отсутствие ранжирования В. Возможность упорядочивать документы по степени предполагаемой близости к запросу С. Работа только с карточными каталогами Д. Обязательное использование только тезаурусов	В	ТК
49.	Какое утверждение верно? А. Все модели поиска выдают результаты одинаково В. Разные модели по-разному определяют соответствие документа запросу С. Булева модель всегда лучше остальных Д. Векторная модель не использует термы	В	ТК
50.	Что означает частота термина в документе? А. Сколько документов есть в коллекции В. Сколько раз терм встретился в данном документе С. Сколько авторов у статьи Д. Сколько языков поддерживает система	В	ТК
51.	Почему редкие содержательные слова часто получают больший вес? А. Потому что они длиннее В. Потому что лучше различают документы по содержанию С. Потому что они всегда стоят в заголовке Д. Потому что они служебные	В	ТК
52.	Что характерно для частотных служебных слов? А. Обычно они слабо различают документы В. Они всегда определяют тему текста С. Они встречаются только в научных статьях Д. Они всегда имеют высокий поисковый вес	А	ТК
53.	Что такое tf-idf на базовом уровне? А. Мера, учитывающая частоту термина в документе и его редкость в коллекции В. Метод ручного индексирования С. Вид предметной рубрики Д. Формат библиографического описания	А	ТК
54.	Что показывает idf? А. Длину документа В. Редкость термина в коллекции С. Число абзацев Д. Порядок слов в фразе	В	ТК
55.	Что чаще всего происходит со словом, встречающимся почти во всех документах коллекции? А. Его значимость как различительного признака обычно снижается В. Его вес автоматически становится максимальным С. Оно становится дескриптором Д. Оно перестает быть токеном	А	ТК

56.	Что такое вес терма? A. Масса печатного издания B. Числовой показатель значимости терма C. Объем PDF-файла D. Количество букв в слове	B	TK
57.	Как частотные характеристики влияют на ранжирование? A. Они не влияют на него B. Они помогают оценить значимость совпавших термов C. Они нужны только для словарей D. Они определяют только дату документа	B	TK
58.	Зачем нужна нормализация длины документа? A. Чтобы длинные документы не получали необоснованное преимущество только из-за большего числа слов B. Чтобы сокращать тексты вручную C. Чтобы удалить метаданные D. Чтобы система работала только с короткими статьями	A	TK
59.	Что из перечисленного является примером стоп-слова? A. Уникальный термин предметной области B. Очень частое служебное слово C. DOI статьи D. Имя редкого автора	B	TK
60.	Что значит дискриминативная сила терма? A. Его способность различать документы по содержанию B. Его длина в символах C. Его место в алфавите D. Его принадлежность к глаголам	A	TK
61.	Что влияет на порядок выдачи результатов? A. Только случайный выбор системы B. Частота терма, его редкость, вес совпадений и другие факторы ранжирования C. Только дата входа пользователя D. Только число открытых окон браузера	B	TK
62.	Почему слово с высокой tf не всегда оказывается самым важным? A. Потому что высокая частота может быть у малосодержательного слова B. Потому что система игнорирует частоту C. Потому что такие слова всегда удаляются D. Потому что важны только метаданные	A	TK
63.	Что такое информативный терм? A. Термин, который хорошо помогает отличать документы друг от друга B. Любое слово длиннее пяти букв C. Только имя автора D. Только термин из словаря	A	TK
64.	Что из перечисленного лучше всего объясняет смысл статистического ранжирования?	A	TK

	А. Документы упорядочиваются по количественным признакам, связанным с запросом и текстом В. Документы сортируются только по алфавиту С. Система выбирает случайную выдачу Д. Документы упорядочиваются только по цвету интерфейса		
65.	Что учитывает tf-idf? А. Только частоту терма в запросе В. Только число документов в коллекции С. И частоту терма в документе, и его распространенность в коллекции Д. Только длину документа	С	ТК
66.	Что такое язык запросов? А. Естественный язык автора документа В. Совокупность правил и средств построения поисковых запросов С. Только язык интерфейса системы Д. Любой разговорный стиль пользователя	В	ТК
67.	Фразовый поиск нужен для: А. Поиска слов в любом порядке В. Поиска точной последовательности слов С. Поиска только по году издания Д. Поиска только по одному терму	В	ТК
68.	Что обычно используется для задания фразового поиска? А. Скобки квадратные В. Кавычки С. Двоеточие Д. Точка с запятой	В	ТК
69.	Поиск по полям означает: А. Поиск только в бумажных каталогах В. Поиск в определенном элементе описания документа, например в поле автора С. Поиск только по иллюстрациям Д. Поиск только на первой странице	В	ТК
70.	Что такое усечение? А. Поиск по части слова для нахождения разных его продолжений В. Исключение документа из базы С. Удаление конца статьи Д. Сокращение аннотации вручную	А	ТК
71.	Что такое маскирование в запросе? А. Скрытие результатов от пользователя В. Использование специальных символов вместо части букв С. Поиск только по метаданным Д. Исключение служебных слов	В	ТК
72.	Фасетный поиск — это: А. Поиск только по одному слову В. Уточнение результатов через выбор категорий и фильтров	В	ТК

	С. Поиск только в полном тексте D. Ручное индексирование		
73.	Что такое фильтр в поисковой системе? A. Средство ограничения результатов по заданным параметрам B. Любой термин из запроса C. Только логический оператор D. Только способ ранжирования	A	TK
74.	Какой пример иллюстрирует поиск по полям? A. автор: Иванов B. "информационный поиск" C. язык* D. термин OR понятие	A	TK
75.	Что означает уточнение запроса? A. Замена научной темы на другую B. Сужение и конкретизация запроса для повышения точности C. Полный отказ от операторов D. Удаление всех фильтров	B	TK
76.	Что означает расширение запроса? A. Добавление новых терминов, синонимов и вариантов для повышения полноты B. Удаление части запроса C. Поиск только по заглавию D. Сортировку результатов по дате	A	TK
77.	Почему важно учитывать особенности синтаксиса разных систем? A. Потому что все системы понимают запросы совершенно одинаково B. Потому что одинаково записанный запрос в разных системах может работать по-разному C. Потому что синтаксис существует только в каталогах D. Потому что это важно только программистам	B	TK
78.	Какой запрос скорее всего будет слишком широким? A. "когнитивная лингвистика" AND metaphor B. language C. автор:Вежбицкая D. "information retrieval" AND corpus AND discourse	B	TK
79.	Какой запрос скорее всего будет слишком узким? A. язык B. дискурс OR текст C. "когнитивная метафора в англоязычном медицинском дискурсе 2018" D. корпус	C	TK
80.	Что из перечисленного лучше всего помогает после неудачного поиска? A. Итеративная модификация запроса B. Полный отказ от анализа результатов C. Использование только одного и того же запроса много раз без	A	TK

	изменений D. Игнорирование синтаксиса системы		
--	--	--	--

Раздел дисциплины (тема) 4: Использование цифровых, программных и интеллектуальных средств в информационном поиске. Применение модуля re и регулярных выражений для поиска по текстовым шаблонам

Уровень применения: Текущий контроль

Вид оценочного средства: Практические задачи

№	Вопрос и варианты ответа	Эталон ответа	Уровень применения
1.	Найти слово data только как отдельное слово Задание: Составьте регулярное выражение, которое находит data только как самостоятельное слово.	Ответ: \bdata\b Комментарий: \b обозначает границу слова. Такое выражение найдет data, но не найдет database.	ТК
2.	Найти все числа в тексте Задание: Составьте регулярное выражение для поиска целых чисел.	Ответ: \d+ Комментарий: \d — цифра, + — одна или более цифр подряд.	ТК
3.	Найти четырехзначные годы Задание: Найдите в тексте все годы вида 1998, 2024, 2010.	Ответ: \b\d{4}\b Комментарий: \d{4} означает ровно 4 цифры	ТК
4.	Найти слова, начинающиеся с заглавной буквы Задание: Найдите слова, начинающиеся с большой латинской буквы.	Ответ: \b[A-Z][a-z]*\b Комментарий: [A-Z] — первая заглавная буква, [a-z]* — далее ноль или более строчных букв.	ТК
5.	Найти адреса электронной почты Задание Составьте регулярное выражение для поиска email-адресов.	Ответ: \b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,}\b Комментарий: Это базовый шаблон для большинства email-адресов.	ТК
6.	Найти слова длиной ровно 5 букв Задание: Найдите в тексте слова, содержащие ровно 5 букв.	Ответ: \b[a-zA-Z]{5}\b Комментарий: {5} задает точное количество символов.	ТК
7.	Найти все слова, оканчивающиеся на ing Задание: Составьте регулярное выражение для поиска английских слов с окончанием ing.	Ответ: \b[a-zA-Z]+ing\b Комментарий: Подходит для слов типа reading, writing, processing.	ТК
8.	Найти повторяющиеся пробелы Задание: Составьте регулярное выражение, которое находит два и более пробела подряд.	Ответ: {2,} Комментарий: Это полезно для очистки текста	ТК
9.	Найти знаки препинания Задание: Найдите в тексте точки, запятые,	Ответ: [.,:;! ?] Комментарий: Квадратные	ТК

	двоеточия, точки с запятой, вопросительные и восклицательные знаки.	скобки обозначают набор допустимых символов.	
10.	Найти слова, содержащие только строчные буквы Задание: Составьте выражение для поиска слов, записанных только строчными латинскими буквами.	Ответ: \b[a-z]+\b	ТК
11.	Найти аббревиатуры из заглавных букв Задание: Найдите аббревиатуры типа NLP, IR, AI, USA.	Ответ: \b[A-Z]{2,}\b Комментарий: Такое выражение ищет последовательности из двух и более заглавных букв.	ТК
12.	Найти слова с дефисом Задание: Найдите слова типа well-known, text-based, problem-solving.	Ответ: \b[a-zA-Z]+-[a-zA-Z]+\b Комментарий: Для нескольких дефисов можно использовать расширенный вариант: \b[a-zA-Z]+(?:-[a-zA-Z]+)+\b	ТК
13.	Найти строки, начинающиеся с цифры Задание: Составьте регулярное выражение для поиска строк, начинающихся с цифры.	Ответ: ^\d Комментарий: ^ — начало строки.	ТК
14.	Найти строки, заканчивающиеся точкой Задание: Составьте регулярное выражение для поиска строк, оканчивающихся точкой.	Ответ: \.\$ Комментарий: \$ — конец строки.	ТК
15.	Найти слова colour и color Задание: Составьте одно регулярное выражение, которое найдет британский и американский варианты написания.	Ответ: \bcolou?r\b Комментарий: u? означает, что буква u может присутствовать или отсутствовать.	ТК
16.	1. Найдите все годы из 4 цифр. 2. Найдите все email-адреса. 3. Найдите все слова, начинающиеся с заглавной буквы. 4. Найдите все даты формата dd.mm.yyyy. 5. Найдите все URL. 6. Найдите все аббревиатуры из 2 и более заглавных букв. 7. Найдите слова с дефисом. 8. Найдите повторяющиеся пробелы. 9. Найдите двойные слова. 10. Найдите слово search только как отдельное слово.	1. \b\d{4}\b 2. \b[A-Za-z0-9._%+~]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,}\b 3. \b[A-Z][a-z]*\b 4. \b\d{2}\.\d{2}\.\d{4}\b 5. https?://S+ 6. \b[A-Z]{2,}\b 7. \b[a-zA-Z]+(?:-[a-zA-Z]+)+\b 8. {2,} 9. \b(\w+)\s+\1\b 10. \bsearch\b	ТК

Раздел дисциплины (тема) 4: Использование цифровых, программных и интеллектуальных средств в информационном поиске.

Вид оценочного средства: Устное выступление с презентацией по теме дисциплины

Уровень применения: Текущий контроль

№	Вопрос и варианты ответа	Эталон ответа	Уровень применения
1. 1	История развития и основные этапы эволюции информационного поиска	Кратко осветить становление информационного поиска от библиотечных каталогов и карточных систем до компьютерных баз данных и современных интернет-поисковиков. Показать этапы развития: ручной поиск, автоматизация каталогов, появление полнотекстового поиска, веб-поиск, интеллектуальный поиск на основе ИИ. Отдельно отметить, как изменялись запросы пользователей и методы обработки документов.	ТК
2.	Классификация поисковых систем (Яндекс, Google, Bing, DuckDuckGo и др.)	Кратко осветить становление информационного поиска от библиотечных каталогов и карточных систем до компьютерных баз данных и современных интернет-поисковиков. Показать этапы развития: ручной поиск, автоматизация каталогов, появление полнотекстового поиска, веб-поиск, интеллектуальный поиск на основе ИИ. Отдельно отметить, как изменялись запросы пользователей и методы обработки документов.	ТК
3.	Индексирование данных: основные принципы и используемые алгоритмы	Объяснить, что такое индексирование и зачем оно необходимо для ускорения поиска. Рассмотреть инвертированный индекс, словари терминов, хранение частот и позиций слов, роль токенизации и нормализации. Упомянуть базовые алгоритмические идеи, лежащие в основе полнотекстового поиска.	ТК
4.	Модели информационного поиска: булева, векторная, вероятностная, латентно-семантический анализ	Дать обзор основных моделей информационного поиска и показать, как каждая из них представляет документы и запросы. Сравнить булев поиск как точное логическое сопоставление, векторную модель как поиск по степени близости, вероятностную модель как оценку релевантности и LSA как работу со скрытыми семантическими связями. Подчеркнуть сильные и слабые стороны каждой модели.	ТК
5.	Поиск информации в интернет-сервисах и специализация поисковых систем	Показать, что поиск осуществляется не только в классических поисковиках, но и внутри маркетплейсов, видеохостингов, научных платформ, социальных сетей и цифровых библиотек. Объяснить, как меняются алгоритмы поиска в зависимости от типа контента. Рассмотреть преимущества	ТК

		специализированных сервисов для решения конкретных задач.	
6.	Поиск по ключевым словам и нечеткие методы поиска: алгоритмы Дамерау-Левенштейна, Хемминга, Джаро-Винклера	Объяснить различие между точным поиском по ключевым словам и нечетким поиском. Рассмотреть, как работают алгоритмы измерения сходства строк и зачем они нужны при опечатках, вариантах написания и неполных совпадениях. Показать практическое применение этих методов в поисковых системах и системах исправления запросов.	ТК
7.	Роль автоматической обработки текста в информационном поиске (NLP, анализ текста, выделение сущностей)	Описать, как методы NLP помогают системе понимать текст глубже, чем простое совпадение слов. Рассмотреть токенизацию, лемматизацию, определение частей речи, извлечение именованных сущностей, анализ контекста. Показать, как это повышает качество поиска и релевантность выдачи.	ТК
8.	Кластеризация и ранжирование документов на основе запросов пользователя и их релевантности	Объяснить, что документы нужно не только находить, но и упорядочивать. Рассмотреть ранжирование как процесс определения наиболее полезных результатов и кластеризацию как способ группировки документов по темам. Показать, как эти методы делают выдачу более понятной и удобной для пользователя.	ТК
9.	Поиск и фильтрация научной информации: базы данных, индексные системы, библиографические сервис	Рассмотреть основные инструменты научного поиска: библиографические базы данных, научные поисковики, системы цитирования, менеджеры литературы. Показать, как фильтровать результаты по тематике, дате, типу публикации, индексу цитирования и ключевым словам. Отдельно подчеркнуть важность оценки качества источника.	ТК
10.	Обзор современных электронных корпусов текстов и их возможностей	Дать представление о корпусах как о специальных коллекциях текстов с инструментами поиска и анализа. Рассмотреть типы корпусов: национальные, параллельные, специализированные, учебные. Показать, какие задачи можно решать с их помощью: поиск словоупотреблений, контекстов, частот, грамматических конструкций.	ТК
11.	Поиск в больших данных и применении технологий искусственного	Показать, как задачи поиска усложняются при работе с огромными объемами разнородных данных. Рассмотреть роль распределенных систем, интеллектуального анализа, машинного обучения и нейросетевых методов. Объяснить,	ТК

	интеллекта.	как ИИ помогает находить связи, прогнозировать интересы пользователя и улучшать релевантность результатов.	
12.	Персонализация поиска: как системы учитывают предпочтения пользователя.	Объяснить, что современные системы подстраивают выдачу под конкретного человека. Рассмотреть учет истории поиска, местоположения, языка, предыдущих кликов и поведения пользователя. Показать преимущества персонализации и возможные риски, связанные с «информационным пузырем».	ТК
13.	Конфиденциальность и безопасность при поиске информации в интернете.	Осветить риски, возникающие при использовании поисковых систем: сбор данных, отслеживание поведения, утечки информации, рекламный профайлинг. Рассмотреть способы защиты: приватные поисковики, режимы анонимности, безопасные соединения, настройки конфиденциальности. Показать, почему цифровая безопасность важна даже при повседневном поиске.	ТК
14.	Обзор современных инструментов для создания эффективных поисковых запросов (практические лайфхаки).	Рассмотреть инструменты и приемы, помогающие искать точнее: кавычки, логические операторы, исключение слов, поиск по сайту, фильтры по типу файла и дате. Показать, как небольшое изменение формулировки запроса улучшает результат. Можно дополнить тему практическими примерами удачных и неудачных запросов.	ТК
15.	Современные поисковые системы: обзор, сравнение, принципы работы и перспективы развития.	Сделать обобщающий обзор крупнейших поисковых систем и сравнить их по качеству поиска, скорости, персонализации, дополнительным сервисам и подходам к конфиденциальности. Объяснить базовые принципы работы: сканирование, индексирование, ранжирование. В завершение обозначить возможные направления развития — семантический поиск, генеративный ИИ, мультимодальность.	ТК
16.	Эффективные стратегии поиска научной информации: практические рекомендации и лайфхаки	Показать, как правильно начинать научный поиск: выделять ключевые понятия, подбирать синонимы, использовать расширенный поиск и операторы. Рассмотреть стратегии работы с библиографией, цитированиями и тематическими подборками. Сделать акцент на практической стороне: как сократить время поиска и повысить качество найденных	ТК

		источников.	
17.	Алгоритмы ранжирования и методы оценки релевантности поисковой выдачи	Рассмотреть, по каким принципам поисковая система определяет порядок результатов. Упомянуть текстовую релевантность, популярность, ссылочные факторы, поведенческие сигналы и семантическое соответствие запросу. Также показать, какими метриками оценивают качество поиска: precision, recall, F-мера и другие.	ТК
18.	Персонализация поиска и роль искусственного интеллекта в информационном поиске	Раскрыть связь между персонализацией и технологиями ИИ. Показать, как алгоритмы машинного обучения анализируют поведение пользователя и подбирают более подходящие результаты. Обсудить не только преимущества, но и ограничения: предвзятость алгоритмов, снижение разнообразия информации, зависимость от пользовательского профиля.	ТК
19.	Методы поиска и фильтрации данных в больших массивах (BigData)	Рассмотреть основные подходы к поиску в больших объемах данных: индексирование, распределенное хранение, фильтрация, агрегация, работа с потоками данных. Показать, почему традиционные методы поиска не всегда эффективны в BigData. Осветить роль автоматизации и интеллектуальных систем в обработке таких массивов.	ТК
20.	Этика и правовые аспекты в информационном поиске: дезинформация, права пользователя, цифровая безопасность	Обсудить, как поиск связан с вопросами ответственности, достоверности и прав пользователя. Рассмотреть проблему дезинформации, манипулятивной выдачи, авторских прав, права на доступ к информации и права на защиту персональных данных. Показать, что информационный поиск — это не только техническая, но и социально-правовая проблема.	ТК
21.	Инструменты поиска в специализированных областях: медицина, лингвистика, техника	Показать, что в разных научных и профессиональных сферах используются особые поисковые инструменты и базы данных. Рассмотреть примеры специализированных ресурсов, особенности запросов и критерии оценки найденной информации. Подчеркнуть, что эффективность поиска зависит от знания предметной области.	ТК
22.	Поиск мультимедийной информации: технологии поиска	Объяснить, как осуществляется поиск не только текста, но и визуального контента. Рассмотреть поиск по метаданным, распознаванию объектов, изображению-образцу, субтитрам,	ТК

	изображений и видео	аудиодорожке и содержанию видео. Показать, как развиваются технологии компьютерного зрения в задачах поиска.	
23.	Поисковые запросы: искусство формулирования и оптимизации	Сосредоточиться на том, как формулировка запроса влияет на качество результата. Рассмотреть типичные ошибки пользователей, способы уточнения запроса, подбор терминов, использование операторов и пошаговое сужение темы. Эту тему удобно строить на практических примерах и сравнении разных вариантов одного запроса.	ТК
24.	Семантический поиск и поиск по смыслу: современные подходы	Рассмотреть отличие традиционного поиска по ключевым словам от семантического поиска, ориентированного на смысл запроса. Показать, как системы используют контекст, связи между понятиями, векторные представления слов и нейросетевые модели. Объяснить, почему такой поиск лучше справляется со сложными, неполными и естественно сформулированными запросами.	ТК
25.	Оценка качества поисковых систем: точность, полнота и удобство использования	Показать, по каким критериям оценивают эффективность поисковой системы. Рассмотреть понятия точности, полноты, релевантности, скорости ответа и удобства интерфейса. Отдельно отметить, что хорошая поисковая система должна не только находить информацию, но и делать ее поиск быстрым, понятным и удобным для пользователя.	ТК

Оценка	Критерии оценивания
«Удовлетворительно»/ зачтено	Ответ полностью соответствует установленным критериям: демонстрирует глубокое понимание материала, логичность изложения, корректное использование терминологии, опору на изученные источники и способность к аргументированной самостоятельной интерпретации. Ответ в целом удовлетворяет предъявляемым требованиям, однако содержит отдельные недостатки: незначительные погрешности в отборе или анализе материала, отдельные нарушения логики изложения. Вместе с тем, общая логика и аргументация сохраняются.
Оценка «Неудовлетворительно» / не зачтено	Содержание источника понято в основных чертах, однако ответ строится преимущественно на репродуктивном воспроизведении материала с опорой на один источник. Самостоятельный анализ и обобщение отсутствуют.

	Ответ свидетельствует о неспособности студента к самостоятельному анализу материала: терминологический аппарат не используется или применяется некорректно, изложение отличается нелогичностью и неполнотой. Оценка выставляется также в случаях, когда ответ не соответствует поставленному вопросу либо обнаруживает полное непонимание темы.
--	--

ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ

Вид оценочного средства: Тестовое задание

Уровень применения: Промежуточная аттестация

№	Вопрос и варианты ответа	Эталон ответа	Уровень применения
1	Что такое булев поиск? A. Поиск с использованием логических операторов B. Поиск по ключевым словам с учётом синонимов C. Поиск по индексированному каталогу D. Ручной просмотр всех доступных страниц	A	ПА
2	Какой показатель измеряет долю общих результатов между двумя поисковыми выдачами? A. Jaccard similarity B. Precision C. Recall D. F1-метрика	A	ПА
3	Что такое релевантность в информационном поиске? A. Степень соответствия найденного документа информационной потребности и запросу пользователя B. Количество просмотров документа другими пользователями C. Позиция документа в общем индексе поисковой системы D. Размер файла и количество страниц в документе	A	ПА
4	Какой из вариантов означает "поиск по точной фразе"? A. "climate change" B. climate change C. climate OR change D. climate AND chang	A	ПА
5	Что означает термин "информационная потребность"? A. Осознанная необходимость в получении определённых сведений B. Желание прочитать интересную статью C. Запрос в поисковой строке D. Любая найденная информация	A	ПА
6	Что такое «метапоисковая система»? A. Система, которая пересылает запрос сразу нескольким	A	ПА

	поисковикам и агрегирует результаты В. Система для поиска по библиотечным каталогам С. Система, которая анализирует тексты внутри PDF D. Поисковик с искусственным интеллектом		
7	Что показывает коэффициент Жаккара (Jaccardsimilarity)? A. Степень пересечения между двумя множествами B. Сколько раз пользователь кликнул на одну и ту же ссылку C. Разность между двумя списками D. Количество элементов в объединении двух множеств	A	ПА
8	Найдите ошибку: Формула: $J(A, B) = (A \cup B) / (A + B - A \cap B)$. Эта формула — ... A. Почти правильная, но перепутаны числитель и знаменатель B. Правильная C. Полностью неверная D. Используется только для топ-5 результатов	A	ПА
9	Что такое регулярное выражение? A. Последовательность символов, описывающая шаблон поиска в тексте B. Обычное описание документа в поисковике C. Указание на формат документа D. Метод анализа ссылок в Интернете	A	ПА
10	Если у двух пользователей по 10 результатов в выдаче, и у них совпало 5 ссылок, чему равен коэффициент Жаккара? A. 0.33 B. 0.25 C. 0.5 D. 0.5	A	ПА
11	Какое из выражений подойдёт для поиска всех слов, начинающихся на "auto"? A. \bauto\w* B. auto+ C. ^auto D. auto.	A	ПА
12	Выражение: $\backslash b \backslash w \{ 4, 6 \}$ Вопрос: Какие слова будут найдены? A. Все слова из 4–6 символов B. Все слова из 6–8 символов C. Слова, начинающиеся на b D. Слова, содержащие ровно 4 буквы	A	ПА
13	Какое определение лучше всего отражает суть информации в контексте теории информации? A. Снятая неопределённость относительно множества возможных исходов события B. Любые данные, записанные на носителе C. Любая последовательность символов некоторого алфавита	A	ПА

	D. Любая текстовая информация в цифровой форме		
14	Что в первую очередь отличает <i>данные</i> от <i>информации</i>? A. Наличие интерпретации и контекста у информации, в отличие от сырых данных B. Данные всегда представлены в цифровом виде, а информация — только в бумажном C. Данные всегда правдивы, а информация может быть ложной Данные относятся только к числам, а информация — к текстам	A	ПА
15	Какое высказывание лучше всего описывает уровень «знание» в пирамиде Data–Information–Knowledge–Wisdom (DIKW)? A. Информация, усвоенная и применимая на практике B. Любые сырые измерения реального мира C. Мнения и оценки пользователя о полученных данных D. Любая текстовая информация, хранящаяся в базе данных	A	ПА
16	Чем отличается информационная потребность от поискового запроса? A. Потребность — внутренняя нужда в информации, а запрос — формализованное выражение этой потребности в системе B. Потребность всегда выражается в виде ключевых слов, а запрос — в виде вопросительного предложения C. Потребность существует только у исследователей, а запрос — у обычных пользователей D. Между ними нет различий: это два названия одного и того же	A	ПА
17	Какой тип поисковой задачи описывает запрос: «Получить численность населения города Москва на 2025 год»? A. Фактографический поиск B. Библиографический поиск C. Документальный поиск D. Аналитический поиск	A	ПА
18	Что такое <i>полнота</i> (recall) в оценке качества информационного поиска? A. Доля найденных релевантных документов от общего числа релевантных документов в коллекции B. Доля релевантных документов среди всех найденных системой документов C. Степень субъективного удовлетворения пользователя результатами D. Общее количество документов в индексе поисковой системы	A	ПА
19	Что такое <i>точность</i> (precision) поисковой системы? A. Доля релевантных документов среди всех документов, выданных системой B. Доля найденных документов от общего числа документов в коллекции C. Способность системы ранжировать документы по дате публикации	A	ПА

	D. Способность системы находить только научные статьи		
20	Что такое <i>релевантность</i> в информационном поиске? A. Степень соответствия найденной информации информационной потребности пользователя B. Количество документов, найденных по одному запросу C. Мера статистической связи между двумя терминами D. Любой документ, содержащий искомое слово	A	ПА
21	Какой пример лучше всего иллюстрирует <i>структурированное</i> хранилище данных? A. Реляционная база данных студентов с полями «Имя», «Фамилия», «Год рождения»⁷ B. Архив сканированных книг в формате изображений C. Набор PDF-файлов статей в папке на компьютере D. Лента постов в социальной сети	A	ПА
22	Какой пример соответствует <i>нестандартизированному</i> (текстовому) хранилищу? A. Коллекция новостных статей в виде сплошных текстов без заранее заданных полей B. База данных с жёстко заданной схемой полей C. Таблица в Excel с строго определёнными столбцами D. Каталог библиотеки в формате MARC с полями автора и года	A	ПА
23	Какой вид поиска лучше всего описывает задачу: «Собрать научные статьи о влиянии искусственного интеллекта на развитие естественных языков для написания обзора»? A. Аналитический поиск B. Библиографический поиск C. Фактографический поиск D. Поиск по автору	A	ПА
24	Что характеризует <i>процедурную</i> информацию? A. Описание того, как выполнять определённые действия или процедуры B. Свод законов и нормативных актов C. Философские рассуждения и гипотезы D. Статистические данные наблюдений	A	ПА
25	Текстовый файл содержит 2500 символов. Сколько это примерно в килобайтах (КБ)? (1 символ = 1 байт; 1 КБ = 1024 байта) A. Около 2,44 КБ B. Ровно 2 КБ C. Около 25 КБ D. Около 0,24 КБ	A	ПА
26	Пользователь вводит в поисковую строку «emojilinguisticanalysis». Что это с точки зрения терминологии информационного поиска? A. Поисковый запрос B. Информационная потребность C. Документ	A	ПА

	D. Информационный ресурс		
27	Какой из примеров показывает <i>фактографический</i> поиск в гуманитарной сфере? A. Узнать год выхода первого издания книги “1984” Дж. Оруэлла B. Найти все статьи о метафорах в политическом дискурсе C. Найти тексты, где часто встречается слово <i>fakenews</i> D. Подобрать монографии по когнитивной лингвистике	A	ПА
28	Что такое предобработка текста (textpreprocessing) в задачах автоматической обработки текста? A. Этап очистки, нормализации и подготовки текста к дальнейшему анализу и машинному обучению B. Процесс ручного редактирования текста перед публикацией C. Только удаление знаков препинания и пробелов D. Исключительно перевод текста с одного языка на другой	A	ПА
29	Какова главная цель предобработки текста в информационном поиске и NLP? A. Привести текст к однородной, структурированной и анализируемой форме B. Сделать текст как можно длиннее для обучения моделей C. Усложнить структуру текста, добавив больше символов D. Сохранить максимальное количество «сырого» шума	A	ПА
30	Что относится к основным целям препроцессинга? A. Удаление шума, унификация текста, снижение размерности признакового пространства B. Добавление в текст новых слов для расширения словаря C. Преобразование текста только в аудиоформат D. Шифрование текста для защиты от анализа	A	ПА
31	Как лучше всего охарактеризовать «шум» в тексте? A. Элементы, не несущие полезной смысловой нагрузки и мешающие анализу B. Только орфографические ошибки C. Любые существительные в тексте D. Только эмодзи и смайлики	A	ПА
32	Почему важно удалять HTML-теги и разметку при предобработке? A. Они описывают оформление, а не смысл, и мешают корректному анализу содержания B. Потому что HTML-теги запрещены в любых текстах C. Потому что они уменьшают объём файла D. Чтобы сделать текст более привлекательным для читателя	A	ПА
33	Зачем приводить текст к одному регистру (обычно к нижнему) на этапе нормализации? A. Чтобы одинаковые по смыслу слова с разным написанием не считались разными токенами B. Чтобы текст выглядел менее официально C. Чтобы отличать имена собственные от нарицательных	A	ПА

	D. Чтобы сократить количество предложений		
34	Что такое токенизация (tokenization) в контексте предобработки текста? A. Разбиение текста на минимальные единицы анализа (токены — слова, предложения и т.п.) B. Перевод текста в двоичный код C. Сжатие текста без потерь D. Замена всех слов в тексте на их синонимы	A	ПА
35	Что характеризует стоп-слова (stop-words)? A. Частотные, малоинформативные слова, часто удаляемые для снижения шума B. Всегда ключевые слова текста C. Слова, которые встречаются только один раз во всём корпусе D. Исключительно междометия и частицы	A	ПА
36	В чём главное отличие лемматизации от стемминга? A. Лемматизация учитывает морфологию и приводит слово к нормальной форме, стемминг просто усекает до основы B. Лемматизация всегда даёт более короткую форму, чем стемминг C. Стемминг полностью совпадает с лемматизацией D. Лемматизация используется только для английского, стемминг — только для русского	A	ПА
37	Какой из ниже перечисленных элементов чаще всего рассматривается как «шум» и подлежит удалению при очистке текста (если не нужен по задаче)? A. HTML-теги, служебный код, лишние пробелы B. Содержательные ключевые слова текста C. Названия терминов предметной области D. Научные термины и понятия	A	ПА
38	Почему препроцессинг помогает снизить размерность признакового пространства? A. Объединяет различные формы одного слова, удаляет стоп-слова и редкие элементы B. Увеличивает количество уникальных токенов C. Добавляет новые искусственные слова D. Дублирует каждое слово несколько раз	A	ПА
39	Какова основная цель систем информационного поиска? A. Найти среди множества документов те, которые релевантны информационной потребности пользователя B. Хранить как можно больше данных без доступа к ним C. Обеспечить быстрый доступ к случайным документам D. Заменить библиотеки и архивы как институты	A	ПА
40	С чем связаны истоки информационного поиска как области знаний? A. С развитием библиотечного дела, документалистики и автоматизации каталогов	A	ПА

	В. С появлением персональных компьютеров С. С изобретением электронной почты Д. С развитием социальных сетей		
41	Какое определение лучше всего описывает поисковую систему общего назначения (например, Яндекс, Google, Bing)? А. Система, индексирующая широкий спектр веб-страниц различных тематик для массового пользователя В. Система, работающая с научными публикациями С. Система, предназначенная для поиска по локальному диску Д. Система, создающая документы и базы данных вместо их поиска	А	ПА
42	Что отличает специализированную (вертикальную) поисковую систему от системы общего назначения? А. Она ориентирована на узкую предметную область (например, научные статьи, вакансии, медицину) В. Она всегда работает только офлайн С. Она не имеет интерфейса для пользователя Д. Она индексирует только изображения и игнорирует тексты	А	ПА
43	Что такое метапоисковая система? А. Система, которая пересылает запрос в несколько других поисковых систем и агрегирует их результаты В. Система, которая индексирует только метаданные самих поисковых систем С. Система, работающая только внутри одной локальной сети Д. Система, которая не показывает результаты поиска пользователю	А	ПА
44	Что такое индекс в информационном поиске? А. Специальная структура данных, позволяющая ускорить поиск документов по терминам В. Список всех URL в Интернете в алфавитном порядке С. Любая таблица в электронной таблице (Excel) Д. Только алфавитный список авторов книг в библиотеке	А	ПА
45	Что такое обратный индекс (invertedindex)? А. Структура, в которой для каждого термина хранится список документов, где он встречается В. Таблица, где для каждого документа перечислены все его термины С. Список ссылок между сайтами без указания термов Д. Индекс, отсортированный по дате, а не по словам	А	ПА
46	Какой из пунктов является одной из задач индексирования данных? А. Выделить термины и создать структуры, обеспечивающие быстрый доступ к документам по этим терминам В. Удалить все документы, в которых встречаются редкие слова С. Хранить документы только в исходном виде без дополнительной обработки Д. Заменить все слова в документе их номерами страниц	А	ПА

47	В чём состоит основная идея векторной модели информационного поиска? А. Документы и запрос представляются как векторы весов термов и сравниваются с помощью меры сходства В. Документы описываются только логическими выражениями С. Для каждого документа моделируется вероятность появления запроса Д. Документы группируются исключительно по авторам и годам	А	ПА
48	Чем характеризуется вероятностная модель информационного поиска? А. Пытается определить вероятность того, что документ релевантен запросу, и ранжирует документы по этой вероятности В. Классифицирует документы только по фиксированным рубрикам С. Использует только совпадение по названию документа Д. Упорядочивает документы по длине текста	А	ПА
49	В чём ключевая идея латентно-семантического анализа (LSA) в информационном поиске? А. Снижение размерности и переход к скрытому пространству тем, где учитываются скрытые связи между терминами и документами В. Полное игнорирование связей между терминами С. Упорядочивание документов исключительно по алфавиту первых слов Д. Удаление всех редких термов как бесполезных	А	ПА
50	Что означает специализация поисковой системы? А. Ориентация на определённый тип контента или предметную область В. Способность работать только в одной стране С. Наличие только одного языка интерфейса Д. Обязательное отсутствие рекламы в результатах	А	ПА
51	Какой тип поисковой системы лучше всего описывает GoogleScholar, PubMed, Scopus? А. Специализированные/вертикальные поисковые системы и библиографические сервисы для научной информации В. Поисковые системы общего назначения С. Метапоисковые системы по всему интернету Д. Корпоративные системы для внутреннего поиска в организации	А	ПА
52	Какова основная идея нечетких методов поиска (fuzzysearch)? А. Нахождение подходящих результатов даже при опечатках, неточном написании или небольших расхождениях в строках В. Выдача пользователю случайных документов С. Полное игнорирование опечаток и ошибок Д. Подбор только документов с точно совпадающими строками	А	ПА

53	Какова роль обработки естественного языка (NLP) в информационном поиске? А. Улучшение понимания текста и запросов: разбор структуры, нормализация слов, определение смысла В. Ускорение работы сети Интернет С. Визуальное оформление результатов поиска Д. Исключительно защита текста от копирования	А	ПА
54	Что такое выделение сущностей (NamedEntityRecognition, NER) в задачах ИП? А. Автоматическое обнаружение имён собственных и значимых объектов (люди, организации, места и т.п.) в тексте В. Автоматический перевод текста на другой язык С. Поиск только по числовым данным Д. Удаление всех имён собственных как «шума»	А	ПА
55	Что такое кластеризация документов? А. Автоматическое группирование документов по сходству содержания без заранее заданных меток В. Разделение документов по алфавиту названий С. Сортировка документов по дате публикации Д. Группировка документов исключительно по длине текста	А	ПА
56	Что означает ранжирование документов в поисковой системе? А. Упорядочивание найденных документов по степени предполагаемой релевантности запросу пользователя В. Упорядочивание документов по размеру файла С. Сортировка документов по алфавиту URL Д. Распределение документов между различными серверами	А	ПА
57	Что представляет собой электронный корпус текстов (корпус языка)? А. Большое структурированное собрание текстов с лингвистической разметкой, предназначенное для анализа языка В. Любой набор случайных текстов в папке С. Только собрание художественной литературы без метаданных Д. База данных только числовых измерений	А	ПА
58	Что отличает поиск в больших данных (BigData) от традиционного поиска? А. Работа с очень большими, разнообразными и быстро обновляющимися наборами данных с использованием распределённых и масштабируемых технологий В. Работа только с небольшими текстовыми файлами С. Отсутствие какой-либо индексации Д. Использование исключительно ручного поиска	А	ПА
59	Какую роль играют технологии искусственного интеллекта в современных поисковых системах?	А	ПА

	A. Улучшение понимания запросов, персонализация, семантический поиск и интеллектуальное ранжирование B. Исключительно ускорение загрузки картинок C. Только защита от вирусов D. Замена всех результатов случайным набором ссылок		
60	Что означает персонализация поиска? A. Учёт истории запросов, кликов, местоположения и предпочтений пользователя для адаптации результатов B. Полный отказ от учёта поведения пользователя C. Выдача одинаковых результатов всем пользователям D. Запрет на использование куки и любого контекста	A	
61	Какой риск связан с персонализацией поиска? A. Формирование «информационного пузыря», когда пользователь видит только узкий круг информации B. Полное отсутствие рекламы C. Необходимость всегда вводить длинные запросы D. Полное исчезновение спама в сети	A	ПА
62	Какой аспект конфиденциальности важен при использовании поисковых систем? A. Сбор и хранение поисковых запросов и логов, которые могут раскрывать интересы и личные данные пользователя B. Только размер шрифта в интерфейсе C. Цветовая схема главной страницы D. Наличие анимации логотипа	A	ПА
63	Какую возможность часто предоставляют поисковые системы для повышения конфиденциальности? A. Режим инкогнито/частного просмотра и использование поисковиков, минимизирующих сбор данных B. Автоматическое сохранение всех запросов публично C. Принудительную привязку всех запросов к имени пользователя D. Запрет на использование шифрования	A	ПА
64	Как можно кратко охарактеризовать современные поисковые системы в целом? A. Сложные программно-аппаратные комплексы, сочетающие индексацию, ранжирование, NLP и машинное обучение B. Небольшие программы, ищущие только по локальным файлам C. Набор статических каталогов без обновления D. Простые таблицы URL без анализа содержимого	A	ПА
65	Пользователь ищет «neuralnetwtok» вместо «neuralnetwork» и всё равно получает релевантные результаты. Какое свойство системы это демонстрирует? A. Использование нечеткого поиска и коррекции опечаток (fuzzymatching) B. Полный отказ от обработки ошибок	A	ПА

	С. Работа только по точному совпадению Д. Игнорирование запроса как ошибочного		
66	Система ИП выделяет в тексте сущности: «Москва», «Сбербанк», «AppleInc.», «COVID-19». Как это может помочь при поиске? А. Позволяет искать документы по конкретным объектам (городам, организациям, заболеваниям) и строить тематические фильтры В. Ухудшает точность поиска С. Используется только для подсчёта длины текста Д. Не влияет на поиск, служит только для оформления	А	ПА
67	По результатам поиска система автоматически группирует документы в кластеры: «Теория ИИ», «Применения ИИ в медицине», «Этика ИИ». Что это даёт пользователю? А. Удобную навигацию по тематическим группам документов В. Случайное перемешивание результатов С. Уменьшение числа найденных документов до одного Д. Полное скрывание части документов	А	ПА
68	Пользователь вводит общий запрос «climatechange», а система предлагает уточнения: «effects», «policy», «healthimpacts». Какой механизм здесь проявляется? А. Анализ запроса и логов, генерация подсказок на основе частотных и связанных запросов В. Случайный подбор слов С. Игнорирование истории запросов других пользователей Д. Полный отказ от персонализации	А	ПА
69	При поиске научной литературы студент использует PubMed и вводит запрос с фильтрами по году, типу статьи и доступу к полному тексту. Как это влияет на результаты? А. Сужает выдачу до более релевантного и актуального подмножества публикаций В. Увеличивает число нерелевантных документов С. Отменяет сортировку по релевантности Д. Показывает только случайно выбранные статьи	А	ПА
70	Исследователь использует национальный корпус языка (например, Российский национальный корпус) для изучения употребления слова «цифровизация». Что он может сделать с помощью корпуса? А. Посмотреть контексты употребления, частоты, сочетаемость и типичные конструкции В. Только скачать тексты без анализа С. Только подсчитать количество букв в слове Д. Только найти перевод слова на другие языки	А	ПА
71	Пользователь заметил, что после ряда запросов о путешествиях ему стали чаще показываться туристические сайты в выдаче. Что это иллюстрирует? А. Персонализацию поиска и адаптацию результатов под интересы	А	ПА

	В. Случайный выбор сайтов С. Жёстко фиксированный набор ссылок Д. Игнорирование истории запросов		
72	Пользователь хочет минимизировать сбор данных о себе. Что он может сделать при работе с поиском? А. Использовать режим инкогнито, поисковики с упором на приватность, отключить персонализацию В. Обязательно входить в аккаунт и сохранять все запросы С. Публиковать свои запросы в социальных сетях Д. Никогда не использовать шифрованные соединения (HTTPS)	А	ПА
73	При индексировании текстов разработчик решает уменьшить размер индекса и ускорить поиск. Какое действие наиболее типично в рамках индексирования? А. Удалить стоп-слова и привести слова к нормальной форме (стемминг/ лемматизация) В. Дублировать все документы дважды С. Удалить из индекса все часто встречающиеся термины, включая ключевые Д. Хранить индекс только в виде печатных таблиц	А	ПА
74	Рассматривается вероятностная модель поиска. Какой принцип ранжирования наиболее ей соответствует? А. Документы упорядочиваются по оценке вероятности релевантности запросу В. Документы упорядочиваются по размеру файла, от меньшего к большему С. Документы упорядочиваются по алфавиту URL Д. Документы сортируются только по дате публикации	А	ПА
75	При использовании LSA запрос «автомобиль» возвращает также документы, где слово «автомобиль» не встречается, но часто встречается слово «машина» в близком контексте. Как это объясняется? А. В скрытом семантическом пространстве термины «автомобиль» и «машина» оказываются близкими по смыслу В. Система случайно добавляет нерелевантные документы С. Система использует только буквенное совпадение терминов Д. Система игнорирует семантическую близость терминов	А	ПА
76	По каким критериям можно отнести GoogleScholar или eLIBRARY к специализированным (вертикальным) системам, а не к общим веб-поисковикам? А. Они ориентированы на поиск научных публикаций и ограничены конкретным типом контента В. Они индексируют вообще весь веб без ограничений С. Они предназначены только для поиска по файлам на локальном компьютере Д. Они используют исключительно булев поиск без ранжирования	А	ПА

77	Что является ключевым принципом эффективного поиска научной информации? А. Чёткое уточнение информационной потребности и использование специализированных научных баз данных В. Ввод максимально длинного свободного текста без фильтров С. Поиск только в популярных социальных сетях Д. Использование лишь первой найденной ссылки из выдачи	А	ПА
78	Какой подход наиболее рационален при поиске литературы для обзора по теме исследования? А. Комбинировать ключевые слова и дескрипторы, использовать фильтры по дате, типу документа и области знаний Ограничиться одним общим словом без уточнений Смотреть только результаты на первой странице Google Брать только статьи, доступные без регистрации	А	ПА
79	Какие две классические метрики оценки качества поисковой выдачи чаще всего используют? А. Точность (precision) и полнота (recall) Скорость загрузки и размер файла Длина URL и число изображений Количество рекламных блоков и цвет интерфейса	А	ПА
80	Какой этический риск особенно актуален в контексте информационного поиска? А. Распространение дезинформации и манипуляция повесткой через ранжирование результатов В. Слишком большой выбор источников С. Слишком короткие текстовые фрагменты в сниппетах Д. Наличие нескольких языков интерфейса	А	ПА
81	Что относится к правам пользователя в сфере цифровой безопасности и поиска? А. Право на защиту персональных данных, информированное согласие и возможность управлять настройками приватности В. Обязанность публиковать историю запросов С. Запрет на шифрование соединений Д. Обязательное предоставление реального имени в запросах	А	ПА
82	Врач-исследователь использует специализированную медицинскую базу (например, PubMed). Чем это лучше общего веб-поиска для научных задач? А. Более строгий отбор источников, метаданные, фильтры по типу исследования и индексирование профильных журналов В. Отсутствие рекламных статей С. Наличие большого выбора статей Д. Поиск только по популярным новостным сайтам	А	ПА
83	Лингвист ищет примеры употребления конструкции в современном	А	ПА

	русском языке. Какой инструмент разумнее выбрать? A. Национальный корпус языка или аналогичный электронный корпус B. Различные форумы C. Печатные газеты и журналы D. Генератор случайных текстов		
84	Исследователь хочет найти систематические обзоры по теме лечения диабета 2 типа. Какой стратегией лучше воспользоваться? A. Использовать специализированные медицинские базы (PubMed и др.), уточнить тип статьи (systematicreview), указать диагноз и ключевые термины B. Ввести в общий поиск «лечение диабета» C. Использовать любую поисковую систему и искать по общим рекомендациям D. Использовать Википедию или любую энциклопедию	A	ПА
85	Пользователь замечает, что после серии запросов о машинном обучении поисковик стал чаще показывать курсы по datascience. Что это иллюстрирует? A. Персонализацию выдачи на основе истории запросов и поведения B. Неконтролируемый случайный выбор результатов C. Случайное совпадение контента D. Ошибку в индексации документов	A	ПА
86	Студент получает в выдаче большое количество сомнительных источников по спорной политической теме. Какой шаг будет этически и методологически правильным? A. Проверить источники на достоверность, сопоставить данные с авторитетными ресурсами и учитывать возможность biases B. Безусловно доверять первым пяти результатам C. Опирается только на анонимные блоги D. Выбрать источники, наиболее соответствующие ключевым словам запроса	A	ПА
87	Пользователь хочет уменьшить влияние персонализации и «пузыря фильтров» при поиске. Какой подход поможет? A. Использовать режим инкогнито, чистить куки, пробовать альтернативные поисковики и анонимизирующие инструменты B. Всегда быть залогиненным с одного устройства C. Отправлять все запросы из одной социальной сети D. Максимально делиться своими данными с сервисом	A	ПА
88	Что такое фильтры в поисковых системах? A. Инструменты, ограничивающие результаты поиска по определённым параметрам (дате, языку, региону и др.) B. Специальные программы, полностью заменяющие работу поисковой системы C. Механизм, который показывает только рекламу и скрывает	A	ПА

	остальные результаты D. Элементы дизайна, отвечающие только за цвет и оформление страницы результатов		
89	Какие критерии лучше всего подходят для оценки надёжности интернет-источника? A. Авторство, дата публикации, официальность и репутация сайта B. Количество иллюстраций, графиков, яркость дизайна, длина текста C. Наличие статистических данных, скорость загрузки, количество баннеров D. Популярность в социальных сетях и число лайков	A	ПА
90	Что лучше всего описывает <i>навигационный</i> тип поиска? A. Пользователь хочет попасть на уже известный ему сайт (например, вводит «vk», «youtube», «гугл почта») B. Пользователь ищет подробную информацию по теме, читает статьи и обзоры C. Пользователь хочет купить товар или скачать файл D. Пользователь ищет точный факт (дату, число, определение)	A	ПА
91	Что лучше всего описывает <i>транзакционный</i> тип поиска? A. Пользователь хочет выполнить действие: купить, скачать, зарегистрироваться, заказать услугу B. Пользователь хочет попасть на официальный сайт организации C. Пользователь хочет узнать точную формулу или дату D. Пользователь просто просматривает случайные страницы без цели	A	ПА
92	Что лучше всего описывает <i>информационный</i> тип поиска? A. Пользователь ищет информацию по теме, чтобы разобраться (читает статьи, обзоры, объяснения) B. Пользователь знает точный сайт и хочет сразу на него перейти C. Пользователь хочет сразу выполнить транзакцию (покупка, заказ) D. Пользователь только проверяет, работает ли интернет-соединение	A	ПА
93	Что лучше всего описывает <i>фактографический</i> тип поиска? A. Пользователь ищет конкретный факт: дату, число, имя, определение (например, «год основания Google») B. Пользователь хочет найти длинную обзорную статью C. Пользователь хочет зарегистрироваться в интернет-магазине D. Пользователь просто изучает структуру сайта без цели получить факт	A	ПА
94	Что такое регулярное выражение в задачах поиска? A. Формальный шаблон для описания множества строк по заданным правилам	A	ПА

	В. Список конкретных слов, разделённых пробелами С. Любая строка текста, введённая пользователем Д. Таблица соответствия слов и их переводов		
95	Для чего чаще всего используют регулярные выражения в информационном поиске и обработке текста? А. Для поиска, замены и выделения подстрок по сложным шаблонам В. Только для подсчёта количества символов в строке С. Исключительно для сортировки строк по алфавиту Д. Только для разбиения текста на предложения	А	ПА
96	Что обозначает метасимвол . (точка) в регулярном выражении? А. Любой одиночный символ, кроме перевода строки (в типичной настройке) В. Только точку как символ С. Только пробел Д. Конец строки	А	ПА
97	Что означает квантификатор +? А. «Одно или более» повторений предыдущего символа или группы В. «Ноль или одно» повторение С. «Ноль или более» повторений Д. «Ровно одно» повторение	А	ПА
98	Что делает квантификатор ? в регулярном выражении? А. Делает предыдущий символ или группу необязательными: «ноль или одно» вхождение В. Запрещает использование предыдущего символа С. Умножает количество вхождений на два Д. Обозначает конец строки	А	ПА
99	Что обозначают квадратные скобки [...] в регулярных выражениях? А. Класс символов: любой один из перечисленных внутри скобок В. Начало и конец строки С. Обязательную группу слов Д. Комментарий, который игнорируется	А	ПА
100	Что значит запись [0-9] в регулярном выражении? А. Любая одна цифра от 0 до 9 В. Строка «0-9» целиком С. Только цифра 0 или цифра 9 Д. Любая последовательность из цифр	А	ПА

№	Вопрос и варианты ответа	Эталон ответа	Уровень примене
---	--------------------------	---------------	-----------------

			ния
1.	Как называются исходные сведения до их осмысления?	Данные	ПА
2.	Как называются осмысленные данные, значимые для пользователя?	Информация	ПА
3.	Как называется носитель зафиксированной информации?	Документ	ПА
4.	Как называется текст, свидетельство или ресурс, используемый в исследовании?	Источник	ПА
5.	Как называется организованная совокупность документов и данных, доступная для использования?	Ресурс	ПА
6.	Как называется система, предназначенная для хранения, обработки и поиска информации?	Система	ПА
7.	Как называется активное выявление информации по запросу пользователя?	Поиск	ПА
8.	Как называется перемещение по уже организованной структуре разделов и рубрик?	Навигация	ПА
9.	Как называется ограничение результатов по заданным параметрам?	Фильтрация	ПА
10.	Как называется словесное выражение информационной потребности пользователя?	Запрос	ПА
11.	Как называется список результатов, который система показывает пользователю?	Выдача	ПА
12.	Как называется степень соответствия результата информационной потребности?	Релевантность	ПА
13.	Как называется показатель доли релевантных документов среди найденных?	Точность	ПА
14.	Как называется показатель доли найденных релевантных документов среди всех релевантных?	Полнота	ПА
15.	Как называется процесс создания поисковых представлений документов?	Индексирование	ПА
16.	Как называется упорядочивание найденных документов по степени соответствия запросу?	Ранжирование	ПА
17.	Как называются сведения, описывающие документ: автор, дата, название, язык?	Метаданные	ПА
18.	Как называется электронное собрание текстов для поиска и языкового анализа?	Корпус	ПА
19.	Как называется специально организованное собрание цифровых объектов?	Коллекция	ПА
20.	Как называется продуманный план организации поиска?	Стратегия	ПА
21.	Как называется изменение исходного запроса для улучшения результатов?	Уточнение	ПА
22.	Как называется средство взаимодействия пользователя с поисковой системой?	Интерфейс	ПА
23.	Как называется совокупность документов, объединённых по какому-либо признаку?	Массив	ПА

24.	Как называется единица поискового представления текста?	Терм	ПА
25.	Как называется отдельный случай появления единицы в тексте?	Вхождение	ПА
26.	Как называется конкретная грамматическая форма слова?	Словоформа	ПА
27.	Как называется базовая словарная форма слова?	Лемма	ПА
28.	Как называется минимальная единица текста, выделяемая системой при обработке?	Токен	ПА
29.	Как называется разбиение текста на минимальные единицы обработки?	Токенизация	ПА
30.	Как называется приведение текста к единообразному виду?	Нормализация	ПА
31.	Как называется приведение слова к его нормальной словарной форме?	Лемматизация	ПА
32.	Как называется усечение слова до основы без полного морфологического анализа?	Стемминг	ПА
33.	Как называется предварительная обработка текста перед анализом и поиском?	Препроцессинг	ПА
34.	Как называются элементы текста, мешающие анализу и не несущие полезной нагрузки?	Шум	ПА
35.	Как называется характеристика буквенного оформления, которую часто унифицируют при обработке текста?	Регистр	ПА
36.	Как называется язык программирования, используемый для автоматизации простых поисковых процедур?	Python	ПА
37.	Как называется файл с командами на Python?	Скрипт	ПА
38.	Как называется подключаемый компонент Python для специальной задачи?	Модуль	ПА
39.	Как называется модуль Python для работы с регулярными выражениями?	re	ПА
40.	Как называется формальное описание множества строк по заданным правилам?	Шаблон	ПА
41.	Как называется часть строки, которую можно искать, выделять или заменять?	Подстрока	ПА
42.	Как называется знак в регулярном выражении, имеющий специальное значение?	Метасимвол	ПА
43.	Как называется элемент регулярного выражения, задающий число повторений?	Квантификатор	ПА
44.	Как называется набор допустимых символов внутри квадратных скобок?	Класс	ПА
45.	Что обозначает запись [0-9] одним словом?	Цифра	ПА
46.	Как называется список контекстов употребления слова в корпусе?	Конкорданс	ПА
47.	Как называется ресурс для уточнения значения,	Словарь	ПА

	употребления и свойств слова?		
48.	Как называется ресурс для поиска библиографического описания книги?	Каталог	ПА
49.	Как одним словом назвать структурированное хранилище сведений о публикациях?	База	ПА
50.	Как называется процесс поиска информации по определённой теме или проблеме?	Тематизация	ПА