

**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«УНИВЕРСИТЕТ УПРАВЛЕНИЯ «ТИСБИ»**

Кафедра информационных технологий

Утверждаю
Зав. кафедрой
О.В.Федорова
Протокол заседания
кафедры № 10
от 06.04.2026

Рабочая программа дисциплины

Наименование дисциплины	Информационный поиск
Направление подготовки	09.03.01 Информатика и вычислительная техника
Направленность (профиль) образовательной программы	Инженерия информационных систем
Год набора	2025, 2026

Составители:

Доцент лаборатории машинного обучения и представления данных
Протасов С.И.
АНО ВО «Университет Иннополис»

Казань

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Разделы рабочей программы:

1. Краткая характеристика дисциплины (модуля).
2. Перечень планируемых результатов обучения.
3. Место дисциплины (модуля) в структуре основной профессиональной образовательной программы высшего образования (ОПОП ВО).
4. Объем, структура и содержание дисциплины (модуля).
5. Учебно-методическое обеспечение.
6. Материально-техническое обеспечение.
7. Оценочные и методические материалы.
8. Дополнительные сведения.

Программа составлена в соответствии с требованиями ОПОП ВО Университета Иннополис по данному направлению подготовки.

Форма обучения	Семестр	Трудоемкость дисциплины в		Контактная работа		Контроль, час.	Самостоятельная работа, час.	Форма промежуточной аттестации (экз./диф. зач./зач.)
		зачетных единиц	академических часах	Лекции, час.	Семинарские (практические) занятия, час.			
очная	7	4	144	30	30	9	75	экзамен

1. КРАТКАЯ ХАРАКТЕРИСТИКА ДИСЦИПЛИНЫ (МОДУЛЯ)

Изучение дисциплины (модуля) «Information Retrieval / Информационный поиск» обеспечивает формирование и развитие компетенций обучающихся в области информационного поиска, их применение для решения различных прикладных задач в рамках профессиональной деятельности. В ходе освоения дисциплины обучающиеся знакомятся с базовыми понятиями теории информационного поиска, основными принципами организации информационно-поисковых систем и алгоритмами аналитико-синтетической переработки документов, учатся осуществлять обоснованный выбор технологии построения информационно-поисковой системы, практически реализовывать основные алгоритмы информационного поиска.

2. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

Целью освоения дисциплины (модуля) является формирование знаний, умений и навыков в рамках профессиональной деятельности, позволяющих успешно работать в избранной отрасли, а также обладать компетенциями, способствующими социальной мобильности, устойчивости и конкурентоспособности обучающегося в условиях современных рыночных отношений. Задачами дисциплины являются изучение основ теории информационного поиска, основных принципов организации информационно-поисковых систем, изучение основных этапов аналитико-синтетической переработки документов, получение навыков владения основными технологиями построения информационно-поисковых систем, технологиями построения тезаурусов и онтологий, а также основными алгоритмами аналитико-синтетической переработки документов.

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

№ п/п	Код компетенции	Наименование компетенции	Индикаторы компетенций	Методы и технологии обучения, способствующие формированию компетенции
1.	ОПК-2	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при	Знать: современные информационные технологии, принципы их работы и методы их использования при решении задач профессиональной деятельности. Уметь: выбирать современные информационные технологии и программные средства, в том числе отечественного производства, для решения задач	Метод дискуссии; Метод контрольных вопросов; Проектная технология; Информационно – коммуникационная технология; Групповая

		решении задач профессиональной деятельности	профессиональной деятельности. Владеть: способами применения необходимых информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности.	технология
2.	ПК-1	Способен разрабатывать требования и проектировать программное обеспечение, создавать и поддерживать среды разработки, автоматизировать и контролировать качество разработки программных решений для информационных систем	Знать: основы структурного, объектно-ориентированного и функционального программирования; стандартные и прикладные библиотеки для выбранного языка программирования; принципы включения стороннего исходного кода (в том числе открытого) в разрабатываемые решения жизненный цикл и этапы разработки ПО; компилируемые и скриптовые языки программирования; языки разметки и представления/хранения данных; архитектуру баз данных (SQL / NoSQL); языки запросов к базам данных (SQL); паттерны проектирования программного кода; архитектуры приложений; общие принципы работы операционных систем; принципы работы микросервисной архитектуры; принципы взаимодействия с API удаленных сервисов (REST, SOAP, PRC); системы управления проектами и задачами (Jira, YouTrack и др.); методики управления IT-проектами; инструменты и методы прототипирования пользовательского интерфейса. Уметь: писать поддерживаемый код на различных языках программирования; определять архитектуру приложений ИС, их масштабируемость; устанавливать и настраивать серверы баз данных; работать с инструментами и методами обработки BigData; эффективно использовать инструменты и системы контроля версий; работать с контейнерами и виртуальными машинами; управлять изменениями в приложениях и автоматизировать их; оценивать временные и материальные затраты на разработку отдельных компонент информационных систем.	Метод дискуссии; Метод контрольных вопросов; Проектная технология; Информационно – коммуникационная технология; Групповая технология

			Владеть: методами и приемами формализации задач, методами и средствами проектирования программного обеспечения, проектирования программных интерфейсов, проектирования баз данных; навыками автоматизированной проверки и заливки кода, сборки компонентов программного решения.	
3.	ПК-2	Способен выполнять работы по созданию (модификации) и сопровождению информационных систем	Знать: архитектуру, устройство и функционирование вычислительных и информационных систем; устройство и функционирование современных информационных систем; современные стандарты информационного взаимодействия систем; программные средства и платформы инфраструктуры информационных технологий организаций; современные подходы и стандарты автоматизации организации (например, CRM, MRP, ERP, ITIL, ITSM); методы выявления требований. Уметь: разрабатывать архитектуру, дизайн и прототипы информационной системы; разрабатывать метрики (количественные показатели) работы информационной системы; настраивать и создавать компоненты анализа программного решения информационной системы; исправлять дефекты и несоответствия в архитектуре и дизайне информационной системы, подтверждать исправление дефектов и несоответствий в коде информационной системы и документации к ней; проверять (верифицировать) архитектуру информационной системы, выполнять параметрическую настройку информационной системы; разрабатывать технологии обмена данными; разрабатывать базы данных информационной системы; осуществлять оптимизацию информационной системы; разрабатывать пользовательскую документацию; оценивать	Метод дискуссии; Метод контрольных вопросов; Проектная технология; Информационно – коммуникационная технология; Групповая технология

			временные и материальные затраты на разработку информационной системы при заданных ресурсах. Владеть: методами оптимизации работы информационной системы и её компонентов; приемами интеграции информационной системы с техническим обеспечением заказчика.	
--	--	--	---	--

Знания: сформированы систематические знания основ теории информационного поиска; основных принципов организации информационно-поисковых систем; основных этапов аналитико-синтетической переработки документов; системы классификации и кодирования информации, в том числе, присвоение кодов документам и элементам справочников; функционирования современных информационных систем; форматов представления документов и их особенностей; основных алгоритмов кластеризации; основ системного анализа; информационно-поисковых языков.

Умения: сформированы умения осуществлять обоснованный выбор технологии построения информационно-поисковой системы; практически реализовывать основные алгоритмы информационного поиска; осуществлять поиск информации и решений на основе экспериментальных действий формулировать требования к информационно-поисковой системе; определять методики морфологических анализаторов; вычислять метрики качества поиска.

Навыки (владения): сформировано владение навыками построения информационно-поисковых систем; построения тезаурусов и онтологий; основными алгоритмами аналитико-синтетической переработки документов; навыками исследования проблем профессиональной деятельности с применением анализа, синтеза и других методов интеллектуальной деятельности; создания каталогов информационно-поисковых систем.

3. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОСНОВНОЙ ПРОФЕССИОНАЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ВЫСШЕГО ОБРАЗОВАНИЯ (ОПОП ВО)

Данная учебная дисциплина (модуль) включена в «Блок 1. Дисциплины (модули)» образовательной программы по направлению подготовки 09.03.01 Информатика и вычислительная техника, направленность (профиль) образовательной программы «Инженерия информационных систем» и относится к обязательной части программы, обеспечивает формирование общепрофессиональных, профессиональных компетенций. Дисциплина (модуль) «Information Retrieval / Информационный поиск» позволяет обучающимся получить знания, умения и навыки для успешной профессиональной деятельности в области программного обеспечения компьютерных вычислительных систем и сетей, автоматизированных систем обработки информации и управления. Дисциплина осваивается студентами при обучении в очной форме на 3 курсе в 7 семестре.

4. ОБЪЕМ, СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объем дисциплины (модуля) составляет:

4 зачетные единицы, 144 академических часа, объем контактной работы — 60 академических часов.

Структура дисциплины (модуля):

№ раздела	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы			
		Очная форма			
		Контактная работа		Контроль, час.	Самостоятельная работа, час.
		Лекции, час.	Семинарские (практические) занятия, час.		
1.	Введение. Информатика и семиотика	10	10	0	25
2.	Общие принципы организации информационно-поисковых систем.	10	10	0	25
3.	Модель информационно-поисковой системы.	10	10	0	25
	Промежуточная аттестация - экзамен			9	
ИТОГО:		30	30	9	75

Содержание дисциплины (модуля):

№ разд ела	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам	Формируемые компетенции
1.	Введение. Информатика и семиотика	Информация. Определение информатики как науки о семантической информации. Проблема восприятия информации.	ОПК-2; ПК-1; ПК-2
2.	Общие принципы организации информационно-поисковых систем	Типы тезаурусов. Форматы представления документов и их особенности. Цели и задачи аналитико-синтетической переработки документов. Индексация документов.	
3.	Модель информационно-поисковой системы	Основные принципы системного анализа. Формулировка требований к информационно-поисковой системе. Выбор шкал для определения меры сходства. Основные алгоритмы кластеризации. Апостериорные правила нахождения весовых коэффициентов.	

5. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

Университет Иннополис обеспечен необходимым комплектом лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства.

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронно-библиотечной системе (электронной библиотеке) и к электронной информационно-образовательной среде Университета (<https://my.university.innopolis.ru>). Электронно-библиотечная система и электронная информационно-образовательная среда обеспечивают возможность доступа, обучающегося из любой точки, в которой имеется доступ к информационно-телекоммуникационной сети «Интернет» (далее — сеть «Интернет»), как на территории Университета, так и вне его.

Также обучающимся обеспечен доступ (удаленный доступ) к современным профессиональным базам данных и информационным справочным системам.

Перечень учебно-методического обеспечения дисциплины (модуля):

Основная литература:

1. Лукашевич, Н. В. Тезаурусы в задачах информационного поиска [Электронный ресурс]: монография / Н. В. Лукашевич. — Электрон. текстовые данные. М.: Московский государственный университет имени М. В. Ломоносова, 2011. — 512 с. — Режим доступа: <http://www.iprbookshop.ru/13346> — ЭБС «IPRbooks».
2. Bruch, S. Foundations of Vector Retrieval. — Publishing House: Springer Cham, 2024. — 185 p. — Режим доступа: <https://link.springer.com/book/10.1007/978-3-031-55182-6> — Полнотекстовая англоязычная БД Springer.

Дополнительная литература:

1. Домашнев, П. А. Методы сортировки и поиска в информационных массивах [Электронный ресурс]: методические указания к лабораторным работам по курсу «Технологии программирования и структуры данных» / П. А. Домашнев, М. Г. Журавлева. — Режим доступа: <http://www.iprbookshop.ru/55642> — ЭБС «IPRbooks».
2. Губарев, В. В. Кибернетика, синергетика, информатика: учебное пособие / В. В. Губарев. — Новосибирск: Сибирский государственный университет телекоммуникаций и информатики, 2009. — 38 с. — Режим доступа: <https://www.iprbookshop.ru/54762.html> — ЭБС «IPRbooks».
3. Dutta, S., Gochhait, S. Information Retrieval in Bioinformatics. A Practical Approach. — Publishing House: Palgrave Macmillan Singapore, 2022. — 157 p. — Режим доступа: <https://link.springer.com/book/10.1007/978-981-19-6506-7> — Полнотекстовая англоязычная БД Springer.

Дистанционная поддержка дисциплины (модуля)

Для дистанционной поддержки дисциплины используется система управления образовательным контентом Moodle, развернутая на портале АНО ВО

«Университет Иннополис». Во время обучения все студенты получают индивидуальные пароли для входа в систему, в которой размещаются: программа курса, контрольные вопросы, задания, темы лекций с презентационными материалами и т. д.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

1. Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

2. Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- специализированным оборудованием, включая демонстрационное оборудование.

Перечень материально-технического оснащения учебных аудиторий конкретизируется приказом директора или уполномоченного им лица ежегодно и размещается на официальном сайте Университета в информационно-коммуникационной сети «Интернет» в подразделе «Материально-техническое обеспечение и оснащенность образовательного процесса. Доступная среда» раздела «Сведения об образовательной организации» по ссылке: <https://innopolis.university/sveden/objects>.

3. Для проведения занятий лекционного и семинарского (практического) типа предлагаются наборы демонстрационного оборудования и учебно-наглядных пособий, обеспечивающие тематические иллюстрации, соответствующие программе дисциплины (модуля).

4. Лаборатории Университета, в том числе Кибер-физическая лаборатория информационной безопасности «Иннокиберполигон» (<https://engineerschool.innopolis.university/innocyberpolygon>), Цифровая лаборатория искусственного интеллекта «InnoDataHub» (<https://engineerschool.innopolis.university/innodatahub>), Цифровая образовательная песочница для разработки передовых систем «DevOps Playground» (<https://engineerschool.innopolis.university/devops>), оснащены лабораторным и специализированным оборудованием. Перечень материально-

технического оснащения и программного обеспечения лабораторий размещается на официальном сайте Университета по ссылке: <https://innopolis.university/sveden/objects>.

5. Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к в электронную информационно-образовательную среду Университета.

6. Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Ссылка на информационный ресурс	Наименование разработки в электронной форме	Доступность
Электронно-библиотечные системы:			
1.	http://www.iprbookshop.ru	Русскоязычная ЭБС на платформе «IPR Smart». Учебники и учебные пособия для университетов издательства «IPRbooks». Учебники и учебные пособия для лиц с ОВЗ	Индивидуальный неограниченный доступ
2.	http://e.lanbook.com	Русскоязычная ЭБС на платформе «Лань». Учебники и учебные пособия для университетов издательства «Лань». Учебники и учебные пособия для лиц с ОВЗ	Индивидуальный неограниченный доступ
Профессиональные базы данных:			
3.	http://search.ebscohost.com	Платформа EBSCOHost. Доступ к англоязычным полнотекстовым книгам	Индивидуальный неограниченный доступ
4.	https://portal.university.innopolis.ru/reading_hall	Электронный каталог научно-технической библиотеки АНО ВО «Университет Иннополис»	Для студентов, зарегистрированных в системе
5.	https://habr.com	База данных для IT-специалистов	Доступ свободный
6.	https://www.sciencedirect.com	База данных ScienceDirect	Доступ свободный
7.	https://elibrary.ru	Научная электронная библиотека	Доступ свободный
8.	https://link.springer.com	Полнотекстовая англоязычная БД Springer	Доступ из локальной сети УИ. Для удаленного доступа требуется регистрация из локальной сети УИ
9.	https://ar.cnki.net	База Данных Academic Reference. База данных полнотекстовых англоязычных ресурсов по всем академическим дисциплинам, опубликованных в Китае	Доступ из локальной сети УИ
10.	https://www.ams.org/journals	База данных англоязычных полнотекстовых журналов AMS Journals (2017–2022 гг.)	Доступ из локальной сети УИ
11.	https://www.dl.begellhouse.com/collections/6764f0021c05bd10.html	База данных англоязычных полнотекстовых журналов Begell	Доступ из локальной сети УИ
12.	https://saemobilus.sae.org/	База данных англоязычных полнотекстовых журналов SAE International SAE eJournals eBooks	Доступ из локальной сети УИ
13.	https://www.worldscientific.com	База данных англоязычных полнотекстовых журналов WorldScientific 2001–2022 гг.	Доступ из локальной сети УИ

14.	https://www.scitation.org/ebooks	AIPP E-Book Collection I + Collection II. Англоязычная База данных полнотекстовых электронных книг.	Доступ из локальной сети УИ
15.	https://ufn.ru/	Автономная некоммерческая организация Редакция журнала "Успехи физических наук". Электронный журнал на русском языке.	Доступ из локальной сети УИ
16.	https://sk.sagepub.com/books/discipline	SAGE Publications Ltd eBook Collections. Англоязычная База данных полнотекстовых электронных книг	Доступ из локальной сети УИ
17.	https://onlinelibrary.wiley.com/	The Wiley Journals Database Англоязычная База данных полнотекстовых электронных журналов	Доступ из локальной сети УИ
18.	https://www.mathnet.ru/	Полнотекстовая коллекция русскоязычных электронных математических журналов	Доступ из локальной сети УИ
19.	https://quantum-electron.lebedev.ru/arhiv/	Электронная версия журнала «Квантовая электроника»	Доступ из локальной сети УИ
20.	http://journals.rcsi.science/	Полнотекстовая коллекция журналов Российской академии наук включает 140 наименований журналов, охватывающих различные научные специальности.	Доступ из локальной сети УИ
21.	https://www.orbit.com/	База данных патентного поиска, объединяющая информацию о более чем 122 миллионах патентных публикаций, полученную из 120 международных патентных ведомств, включая РосПатент, Всемирную организацию интеллектуальной собственности (ВОИС), Европейскую патентную организацию.	Доступ из локальной сети УИ
Информационные справочные системы:			
22.	https://minobrnauki.gov.ru/	Официальный сайт Министерства науки и высшего образования Российской Федерации	Доступ свободный
23.	https://www.edu.ru/	Федеральный портал «Российское образование»	Доступ свободный
24.	http://window.edu.ru/	Информационная система "Единое окно доступа к образовательным ресурсам"	Доступ свободный
25.	http://school-collection.edu.ru/	Единая коллекция цифровых образовательных ресурсов	Доступ свободный
26.	http://fcior.edu.ru/	Федеральный центр информационно-образовательных ресурсов	Доступ свободный

7. Университет Иннополис обеспечен необходимым комплектом лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, в который входят:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое

Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Системы управления обучением:		
Система LMS Moodle	зарубежное	свободно распространяемое
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. МЕТОДИЧЕСКИЕ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

Задания для семинарских (практических) занятий:

№ п/п	Наименование раздела дисциплины (модуля)	Перечень рассматриваемых тем (вопросов)
1.	Введение. Информатика и семиотика	Практическая работа: обсуждение понятия «информация». Практическая работа: Соотношение между понятиями «данные», «информация» и «знания». Практическая работа: Проблема восприятия информации. Практическая работа: Знак, концепт, денотат – треугольник Фреге.
2.	Общие принципы организации информационно-поисковых систем	Практическая работа: Поиск и распространение информации и ГОСТ Практическая работа: Проект Semantic Web: достоинства и недостатки.
3.	Модель информационно-поисковой системы	Практическая работа: Модель информационно-поисковой системы, отвечающая основным системным принципам. Практическая работа: Модели поиска: контекстный, атрибутивный, «по аналогии». Практическая работа: Создание каталогов информационно-поисковых систем.

Текущий контроль успеваемости обучающихся по дисциплине (модулю):

№ п/п	Наименование раздела дисциплины (модуля)	Форма текущего контроля	Материалы текущего контроля
1.	Введение. Информатика и семиотика	Проверка выполнения домашних заданий; Устный / письменный опрос; Защита проекта; Коллоквиум	Задания, в том числе, для групповых проектов: 1. Выберете на http://db.chgk.info/ 10–20 случайных вопросов. Зарегистрируйтесь в поисковой системе, включите запись запросов (если отключали). Историю запросов можно посмотреть и настроить по адресам https://history.google.com/history/, http://nahodki.yandex.ru/. Попробуйте найти ответ на вопрос в интернете (если вы знаете ответ без интернета или по точному совпадению с вопросом сразу находится ответ - переходите к следующему). Не тратьте больше 10 минут на вопрос. Используйте минимум две поисковые машины (для разных вопросов). Оформите лог поиска (в том числе для вопросов, ответ на которые не удалось найти): запросы с временными метками, действия и рассуждения между запросами (запрос - время, "посмотрел на сниппеты, понял, что ничего подходящего нет, новый запрос"; запрос - время; "в документе на позиции 3 - URL - нашел список жен Петра I"; запрос - время). По результатам выполнения задания сформулируйте критерии для оценки качества поисковой системы. Сравните использованные системы. 2. Ознакомьтесь с морфологическими анализаторами (настройки, параметры, форматы входных и выходных данных): Rymorphy2 mystem Яндекса АОТ Алексея Сокирко Поэкспериментируйте с документами разных жанров: художественная литература, научные статьи, сообщения на форумах. Обратите внимание, как обрабатываются редкие слова, слова с опечатками, иностранные слова, имена собственные. Предложите методику оценки морфологического анализатора, сравните по этой методике любые два анализатора. Результаты оформите в виде отчета.
2.	Общие принципы организации информационно-поисковых систем	Проверка выполнения домашних заданий; Устный / письменный опрос; Защита проекта; Коллоквиум	Задания, в том числе, для групповых проектов: 1. Постройте по данным BY.WEB таблицу. Данные, которые должны быть представлены в таблице: Количество документов Средний размер документа в байтах Средний размер документа в байтах после удаления разметки Средняя длина документа в словах Средняя длина слова в буквах Размер коллекции в словах Размер словаря коллекции Словопозиции. 2. Постройте зависимость размера словаря от размера коллекции. Аппроксимируйте прямой в log-log координатах. Нарисуйте график. Постройте частотный словарь коллекции, зависимость частоты от ранга в словаре. Аппроксимируйте прямой в log-log координатах. Нарисуйте график. Проанализируйте результаты. Индексирование коллекции BY.WEB Проиндексировать коллекцию By.Web (см. описание формата)

			одним из поисковых "движков": Terrier Apache Lucene Indri / Galago (Lemur Project) Sphinx В отчете описать предпринятые шаги, использованные настройки, возникшие проблемы и их решение, а также указать время индексирования.
3.	Модель информационно-поисковой системы	Проверка выполнения домашних заданий; Устный / письменный опрос; Защита проекта; Коллоквиум	Задания, в том числе, для групповых проектов: 1. Ознакомьтесь с задачей и форматом данных. Постройте функцию ранжирования (см. ссылки на библиотеки ниже) на основе обучающего множества, примените ее к тестовым данным, загрузите решение. В отчете опишите методы, приведите ваши результаты со страницы рейтинга. 2. Реализуйте вычисление и сохранение значений признаков (features) в индексе для пар "запрос-документ", которые могут быть полезным сигналом для ранжирования. Используйте оценки релевантности пар "запрос-документ" для коллекции. Вычислите значений метрик качества поиска, сравните со стандартным решением.

Контрольные вопросы для подготовки к промежуточной аттестации:

№ п/п	Наименование раздела дисциплины (модуля)	Наименование вопроса
1.	Введение. Информатика и семиотика	Краткая история информатики как науки о семантической информации. Многоуровневая модель информации. Соотношение между понятиями «данные», «информация» и «знания». Семиотика и проблема восприятия информации. Знак, концепт, денотат – треугольник Фреге. Связь между многоуровневой моделью информации и семиотическими моделями История возникновения тезаурусов и онтологий.
2.	Общие принципы организации информационно-поисковых систем	Типы тезаурусов. Многообразие определения онтологий. Типы классификации документов. Цели и задачи аналитико-синтетической переработки документов. Основные этапы аналитико-синтетической переработки документов. Форматы представления документов и их особенности. Информационно-поисковые языки Проект Semantic Web: достоинства и недостатки
3.	Модель информационно-поисковой системы	Сравнение наиболее распространенных классификаторов (УДК, ГРНТИ, MSC). Понятие метаданных. Типы и иерархия метаданных. Стандарт Dublin Core. Особенности аналитико-синтетической переработки электронных документов. Библиометрика: исследование количественных характеристик документов. Понятие об импакт-факторе. Основные библиометрические индексы: индекс цитирования, индекс Хирша и его модификации. Вебометрика: особенности количественного анализа интернет-ресурсов. Применение теории графов в вебометрических исследованиях. Системный анализ. Основные принципы системного анализа. Сформулируйте требования к информационно-поисковой системе. Основные алгоритмы кластеризации.

Вопросы/Задания к промежуточной аттестации в устной/письменной форме:

Выберите правильный вариант ответа.

1. Какая из метрик качества одновременно:

- может быть применена к одной странице результатов поисковой системы (SERP)

- учитывает ранг элементов выдачи

a) полнота (в терминах информационного поиска)

b) discounted cumulative gain (DCG)

c) cumulative gain (CG)

d) аккуратность

e) полнота (в терминах машинного обучения)

f) pFound

g) A/B тест

h) Mean reciprocal rank (MRR)

2. Рассчитайте метрику Precision@K для K=8 если нам даны следующие оценки релевантности: { 1, 0, 1, 0, 1, 1, 1, 1, 0, 0 }.

a) 0.5

b) 0.625

c) 0.75

d) 1.0

3. Какие IP-адреса принадлежат Yandex (отвечают на запросы [*.]yandex.ru)?

Отметьте все. Для ответа используйте любые доступные средства.

a) 5.255.255.77

b) 213.180.193.56

c) 192.34.57.61

d) 77.88.55.77

4. Представьте, что вы сканируете WiFi трафик, чтобы перехватить пароль. Некто установил TCP соединение с 80 портом машины 192.34.57.61. Затем в логах вы видите строку: GET /files/my.password HTTP/1.1\nhost: sprotasov.ru\n\n

Чему равен пароль?

a) HTTP/1.1 404 Not Found

b) HTTP/1.1 200 OK

c) :`(`

d) No such URL :(

e) IDK

f) sprotasov.ru

g) ;)

h) :)

5. Что общего у следующих слов: {fair, unfair, fairness}

a) основа

b) часть речи

c) лемма

6. Для слова "fax" какие из исправления опечаток имеют наивысший коэффициент IoU для биграмма (коэффициент Жаккарда). Не включайте границы слова, биграммы слова fax: {fa, ax}.

- a) **axe**
- b) fox
- c) fix
- d) faximile

7. Слово "чикипарабум" используется в 3 из 15 документах коллекции. Один из этих документов "О е чикипарабум рамамбахара чикипарабум бум", где {e, рамамбахара} — это стоп-слова. Каково значение TF-IDF, если вы используете следующую формулу?

$$\text{TF-IDF} = \text{TF}(t, d) * \text{IDF}(t) = \text{TF}(t, d) * [\log_2((1 + n_{\text{doc}}) / (1 + \text{doc_freq}(d, t)) + 1)]$$

- a) 0
- b) 0.5
- c) 1
- d) **1.5**
- e) 3.0

8. Нейросеть с одним скрытым слоем, которая предсказывает слово, учитывая его близкий контекст, называется:

- a) вложения для языковых моделей
- b) непрерывные скип-граммы
- c) **непрерывный мешок слов**
- d) вектор параграфа — распределённый мешок слов
- e) глубокая структурированная семантическая модель

9. Отметьте все техники, которые могут быть использованы для уменьшения размерности данных в датасете:

- a) **Латентный семантический анализ**
- b) кластеризация k-средних
- c) **Случайные проекции**
- d) кластеризация c-средних

10. Вы проводите латентный семантический анализ с помощью сингулярного разложения документ-терм матрицы, определённой для D=1000 документов и W=1000 слов словаря. Вы нашли низко-ранговое разложение на 2 матрицы с рангом R=10. Матрицы разложения плотные, и вы используете числа с двойной точность (8 байт). Сколько байт вам необходимо для хранения матриц из разложения?

- a) 20000
- b) 40000
- b) 80000
- c) **160000**
- d) 320000

11. Какой из методов кластеризации не гарантирует выпуклую форму кластеров?

- a) **DBSCAN**
- b) Двоичное разбиение пространства (BSP)
- c) k-средних

12. Что из нижеперечисленного не является методом кластеризации?
- a) DBSCAN
 - b) Двоичное разбиение пространства (BSP)
 - c) k-средних
 - d) k ближайших соседей**
13. Что из нижеперечисленного можно считать методом кластеризации, образующим выпуклые кластеры, но эти разделы не создают диаграмму Вороного?
- a) DBSCAN
 - b) Двоичное разбиение пространства (BSP)**
 - c) k-средних
 - d) k ближайших соседей
14. Даже если метод гарантирует выпуклую форму кластера, он не обещает сделать кластеры сбалансированными по количеству элементов.
- a) DBSCAN
 - b) Двоичное разбиение пространства (BSP)
 - c) k-средних**
 - d) k ближайших соседей
15. Вы хотите использовать KD-дерево для поиска ближайших соседей и поиска в диапазоне. В индексе будут храниться $k=17$ -мерные данные. Какие размеры набора данных N подходят для такого индекса? Отметьте все.
- a) 100
 - b) 10 000
 - c) 1 000 000**
 - d) 100 000 000**
16. Отметьте необходимые условия для построения индекса с использованием метризованного графа "мир тесен"?
- a) В узлах элементы из векторного пространства
 - b) В узлах элементы из метрического пространства**
 - c) Граф связный**
 - d) Узлы графа нельзя дополнять
17. Какой из следующих методов лучше всего подходит для извлечения всех кадров с автомобилем "Формулы-1" из видеозаписи во время гонки при наличии одного снимка автомобиля?
- a) вейвлет-хэш
 - b) SIFT**
 - c) текстурные признаки Харалика
 - d) свёрточный автоэнкодер
 - e) признаки Хероха
18. Мы знаем, что ухо человека имеет верхний предел восприятия частот около 20 кГц. Мы знаем, что теорема Котельникова-Найквиста-Шеннона гарантирует, что все частоты вплоть до F Гц точно восстанавливаются для частоты дискретизации $2F$ Гц. Мы знаем, что где-то есть источник звука с частотой выше 20 кГц. Представьте, что для записи мы делаем выборку с частотой 40 кГц. Какое утверждение верно?
- a) Мы заметим потерю качества низких частот при записи.

- b) Мы заметим потерю качества высоких частот при записи.
 - c) **При проигрывании записи может возникнуть низкочастотный шум.**
 - d) Качество восстановленного сигнала будет неотличимо от оригинала.
19. Предложите лучшее дополнение запроса "cookie recipe". Должна увеличиться полнота (больше релевантных документов) без потери точности (не добавляются нерелевантные).
- a) +monster
 - b) +soup
 - c) +Trump
 - d) **+biscuit**
20. Какой термин соответствует субъективному восприятию пользователем того, что результат поиска соответствует информационной потребности пользователя:
- a) аккуратность
 - b) точность
 - c) **релевантность**
 - d) ранг
21. Какова глубина однородного KD-дерева с 1000 элементов?
- a) 5
 - b) **10**
 - c) 17
 - d) 33
 - e) 100
22. Какой тип рекомендательных систем вы можете предложить для нового сервиса без записанной пользовательской активности?
- a) **основанный на содержании**
 - b) коллаборативная фильтрация
 - c) **основанный на правилах**
23. Техника, которая улучшает результаты поиска на основе оценки пользователем первоначальных результатов поиска, называется:
- a) **обратная связь по релевантности (relevance feedback)**
 - b) коллаборативная фильтрация
 - c) расширение запроса
24. Техника, которая улучшает результаты поиска на основе точной настройки запроса пользователя:
- a) обратная связь по релевантности (relevance feedback)
 - b) коллаборативная фильтрация
 - c) **расширение запроса**
25. And approach to recommendations, which uses behavioural information of multiple users is called:
- a) обратная связь по релевантности (relevance feedback)
 - b) **коллаборативная фильтрация**
 - c) расширение запроса
26. Какой из методов тестирования используется для проверки гипотез на реальных пользователях?
- a) Модульное тестирование

- b) Анализ рынка
 - c) A/B-тестирование**
 - d) регрессионное тестирование
27. Идея о том, что распределение слов в текстах может быть аппроксимировано небольшим количеством распределений, каждое из которых описывает специальную область знаний, называется:
- a) доменное моделирование
 - b) тематическое моделирование**
 - c) сингулярное разложение
 - d) линейное разложение
28. Какая архитектура нейронных сетей была разработана специально для информационного поиска:
- a) word2vec
 - b) GloVe
 - c) BERT
 - d) DSSM**
29. Восприятие звука человеком базируется на идее:
- a) частотного разложения**
 - b) высокочастотного семплирования
 - c) понимания низких частот
 - d) понимания высоких частот
30. Минимальный структурный элемент человеческой устной речи — это:
- a) лексема
 - b) токен
 - c) фонема**
 - d) слово
 - e) буква

Критерии оценивания сформированности компетенций

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

В рамках внутренней системы оценки качества образовательной деятельности по программе обучающимся предоставляется возможность оценивания условий, содержания, организации и качества образовательного процесса по данной дисциплине (модулю).

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме экзамена, при этом проводится оценка компетенций, сформированных по дисциплине.

Система оценивания результатов обучения по дисциплине (модулю)

Знания, умения и навыки обучающихся при промежуточной аттестации определяются в следующей форме: «отлично» (А), «хорошо» (В), «удовлетворительно» (С), «неудовлетворительно» (D).

Критерии оценивания результатов обучения по дисциплине (модулю)

Результат обучения по дисциплине (модулю)	Общая характеристика результата обучения по дисциплине (модулю)
А «Отлично»	Обучающийся владеет знаниями предмета в полном объеме рабочей программы, достаточно глубоко осмысливает дисциплину; самостоятельно, в логической последовательности и исчерпывающе отвечает на все вопросы, подчеркивая при этом самое существенное, умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделять в нем главное: устанавливать причинно-следственные связи; четко формирует ответы, свободно читает результаты анализов и других исследований и решает ситуационные задачи повышенной сложности; хорошо знаком с основной литературой и методами исследования большого объема, необходимым для практической деятельности; увязывает теоретические аспекты предмета с задачами практического значения.
В «Хорошо»	Обучающийся владеет знаниями предмета практически в полном объеме рабочей программы; самостоятельно, в логической последовательности и исчерпывающе отвечает на все вопросы, подчеркивая при этом самое существенное, умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделять в нем главное: устанавливать причинно-следственные связи; четко формирует ответы, свободно читает результаты анализов и других исследований и решает ситуационные задачи повышенной сложности; хорошо знаком с основной литературой и методами исследования большого объема, необходимым для практической деятельности; увязывает теоретические аспекты предмета с задачами практического значения.
С «Удовлетворительно»	Обучающийся владеет основным объемом знаний по дисциплине; проявляет затруднения в самостоятельных ответах, оперирует неточными формулировками; в процессе ответов допускаются ошибки по существу вопросов. Обучающийся способен решать лишь наиболее легкие задачи, владеет только обязательным минимумом методов исследований.
D «Неудовлетворительно»	Обучающийся не освоил обязательного минимума знаний предмета, не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.

Методические указания для обучающихся по освоению дисциплины (модуля)

Вид учебных занятий/деятельности	Деятельность обучающегося
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения лекции, выводы, формулировки, обобщения; пометить важные мысли, выделять ключевые слова, термины. Обозначить вопросы, термины или другой материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации, во время семинарского (практического) занятия.
Практическое (семинарское) занятие	При подготовке к семинарскому (практическому) занятию необходимо проработать материалы лекций, основной и дополнительной литературы по заданной теме. На основании обработанной информации постараться сформировать собственное мнение по выносимой на обсуждение тематике. Обосновать его аргументами, сформировать список источников, подтверждающих его. Во время семинарского (практического) занятия активно участвовать в обсуждении вопросов, высказывать аргументированную точку зрения на проблемные вопросы. Приводить примеры из источниковой базы и научной и/или исследовательской литературы.

Устный/письменный опрос	Отвечать, максимально полно, логично и структурировано, на поставленный вопрос. Основная цель – показать всю глубину знаний по конкретной теме или ее части.
Подготовка к промежуточной аттестации	При подготовке к промежуточной аттестации необходимо проработать вопросы по темам, которые рекомендуются для самостоятельной подготовки. При возникновении затруднений с ответами следует ориентироваться на конспекты лекций, семинаров, рекомендуемую литературу, материалы электронных и информационных справочных ресурсов, статей. Если тема вызывает затруднение, четко сформулировать проблемный вопрос и задать его преподавателю.
Практические (лабораторные) занятия	Практические занятия предназначены прежде всего для разбора отдельных сложных положений, тренировки аналитических навыков, а также для развития коммуникационных навыков. Поэтому на практических занятиях необходимо участвовать в тех формах обсуждения материала, которые предлагает преподаватель: отвечать на вопросы преподавателя, дополнять ответы других студентов, приводить примеры, задавать вопросы другим выступающим, обсуждать вопросы и выполнять задания в группах. Работа на практических занятиях подразумевает домашнюю подготовку и активную умственную работу на самом занятии. Работа на практических занятиях в форме устного опроса заключается прежде всего в тренировке навыков применять теоретические положения к самому разнообразному материалу. В ходе практических занятий студенты работают в группах для обсуждения предлагаемых вопросов.
Самостоятельная работа	Самостоятельная работа состоит из следующих частей: 1) чтение учебной, справочной, научной литературы; 2) повторение материала лекций; 3) составление планов устных выступлений; 4) подготовка видеопрезентации. При чтении учебной литературы нужно разграничивать для себя материал на отдельные проблемы, концепции, идеи. Учебную литературу можно найти в электронных библиотечных системах, на которые подписан АНО Университет Иннополис.
Доклад	Публичное, развернутое сообщение по определенной теме или вопросу, основанное на документальных данных. При подготовке доклада рекомендуется использовать разнообразные источники, позволяющие глубже разобраться в теме. Учебную литературу можно найти в электронных библиотечных системах, на которые подписан АНО Университет Иннополис.
Дискуссия	Публичное обсуждение спорного вопроса, проблемы. Каждая сторона должна оппонировать мнению собеседника, аргументируя свою позицию.
Тестирование (устное/письменное)	При подготовке к тестированию необходимо проработать материалы лекций, семинаров, основной и дополнительной литературы по заданной теме. Основная цель тестирования – показать уровень сформированности знаний по конкретной теме или ее части.
Выполнение домашних заданий и групповых проектов	Для выполнения домашних заданий и групповых проектов необходимо получить формулировку задания от преподавателя и убедиться в понимании задания. При выполнении домашних заданий и групповых проектов необходимо проработать материалы лекций, основной и дополнительной литературы по заданной теме.

Для подготовки к занятиям рекомендуется использовать представленные источники в электронных форматах и дополнительную литературу.

8. ДОПОЛНИТЕЛЬНЫЕ СВЕДЕНИЯ

Данная программа и/или ее фрагменты не могут быть использованы другими образовательными организациями без соответствующего разрешения АНО ВО «Университет Иннополис».

