

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Таскаев Сергей Валерьевич
Должность: Ректор
Дата подписания: 12.09.2025 09:43:23
Уникальный программный ключ:
04c19ed8bf98f3b6cb77a486b9a8788b8322323



МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение
ФГБОУ ВО «ЧелГУ»

Версия документа - 1	стр. 1	Первый экземпляр _____	КОПИЯ № _____
----------------------	--------	------------------------	---------------

**Фонд оценочных средств
для промежуточной аттестации
по дисциплине**

**СОВРЕМЕННЫЕ ТЕХНОЛОГИИ ПОИСКА И ОБРАБОТКИ
ИНФОРМАЦИИ**

Направление подготовки

45.04.02 Лингвистика

Направленность (профиль)

Перевод и переводоведение

Присваиваемая квалификация
магистр

Форма обучения
очная

Челябинск 2025 г.

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 2	Первый экземпляр _____	КОПИЯ № _____

Содержание

1. Паспорт фонда оценочных средств
2. Перечень формируемых компетенций
 - 2.1. Компетенции, закрепленные за дисциплиной
3. Содержание оценочных средств по дисциплине
 - 3.1. Виды оценочных средств
 - 3.2. Содержание оценочных средств
4. Порядок проведения и критерии оценивания промежуточной аттестации
 - 4.1. Порядок проведения промежуточной аттестации
 - 4.2. Критерии оценивания промежуточной аттестации по видам оценочных средств
 - 4.3. Результаты промежуточной аттестации и уровни сформированности компетенций

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 3	Первый экземпляр _____	КОПИЯ № _____

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

Направление подготовки 45.04.02 Лингвистика

Направленность (профиль) Перевод и переводоведение

Дисциплина: «Современные технологии поиска и обработки информации»

Семестр изучения: 1

Форма промежуточной аттестации: зачет

Для оценивания результатов обучения по дисциплине используется балльно-рейтинговая система. На основании Методических рекомендаций по использованию балльно-рейтинговой системы оценки результатов обучения студентов в образовательном процессе для направления подготовки 45.04.02 Лингвистика направленности (профиля) Перевод и переводоведение оценка учебных достижений студента по дисциплине может складываться из суммы набранных баллов / оценок за посещаемость, текущий контроль, промежуточную аттестацию, бонусный балл.



2. ПЕРЕЧЕНЬ ФОРМИРУЕМЫХ КОМПЕТЕНЦИЙ

2.1. Компетенции, закреплённые за дисциплиной

Изучение дисциплины «Современные технологии поиска и обработки информации» направлено на формирование следующих компетенций:

Таблица 1

Коды компетенции согласно ФГОС ВО	Содержание компетенций согласно ФГОС ВО	Индикаторы достижения компетенции согласно ОПОП	Перечень планируемых результатов обучения по дисциплине
1	2	3	4
УК-4	Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.2. Демонстрирует умение применять современные коммуникативные технологии для академического и профессионального взаимодействия в ситуации устной и письменной коммуникации, в том числе на иностранном (ых) языке(ах)	Для достижения УК-4.2. знать: сущность и значение информации в развитии современного общества, основные закономерности создания и функционирования информационных процессов; методы и технологии поиска и обработки информации средствами Интернет и офисных приложений Для достижения УК-4.2. уметь: работать с информацией в глобальных компьютерных сетях; применять для академического и профессионального взаимодействия в ситуации устной и письменной коммуникации, в том числе на иностранном(ых) языке(ах) современные информационные технологии поиска и обработки информации Для достижения УК-4.2. владеть: основными методами, способами и средствами поиска, получения, хранения и обработки информации
ПК-4	Способен решать собственные научно-исследовательские и профессиональные задачи на основе анализа исследований в области лингвистики, межкультурной коммуникации и переводоведения	ПК-4.1. Адекватно анализирует современное состояние научных исследований в области лингвистики, межкультурной коммуникации и переводоведения	Для достижения ПК-4.1. знать: методологические принципы поиска, хранения, обработки и анализа информации Для достижения ПК-4.1. уметь: осуществлять поиск, хранение, обработку и анализ информации при проведении научных исследований в области лингвистики, межкультурной коммуникации и переводоведения Для достижения ПК-4.1. владеть: навыками практического использования современных информационных систем и баз данных при проведении научных



МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение
ФГБОУ ВО «ЧелГУ»

Версия документа - 1

стр. 5

Первый экземпляр _____

КОПИЯ № _____

			исследований в области лингвистики, межкультурной коммуникации и переводоведения
--	--	--	--



3. СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

3.1 Виды оценочных средств

Таблица 2

№ п/п	Код компетенции/ планируемые результаты обучения	Контролируемые темы/ разделы	Наименование оценочного средства для текущего контроля	Наименование оценочного средства на промежуточ- ной аттестации/ № задания
1	УК-4 Для достижения УК-4.2. знать: сущность и значение информации в развитии современного общества, основные закономерности создания и функционирования информационных процессов; методы и технологии поиска и обработки информации средствами Интернет и офисных приложений	Определение информации. Виды информации. Источники информации. Системы поиска и обработки информации. История и современное состояние. Информационная грамотность и информационная культура. Защита информации.	Доклад Опрос в устной форме Учебная задача	Опрос Вопрос 2
	Для достижения УК-4.2. уметь: работать с информацией в глобальных компьютерных сетях; применять для академического и профессионального взаимодействия в ситуации устной и письменной коммуникации, в том числе на иностранном(ых) языке(ах) современные информационные технологии поиска и обработки информации	Роль информации при решении проблемы. Проблемы поиска информации. Отбор информации. Критерии отбора информации. Проблемы анализа информации. Алгоритм анализа документальных источников информации. Технологии анализа электронной информации. Контентный анализ.	Доклад Опрос в устной форме Учебная задача	Опрос Вопрос 1, 7
	Для достижения УК-4.2. владеть: основными методами, способами и средствами поиска, получения, хранения и обработки информации	Формы, методы и средства идентификации, поиска и хранения информации. Интернет, типы ресурсов Интернета. Особенности поиска информации в Интернете. Поисковые системы Google; Яндекс; Апорт; AltaVista. Тематические каталоги. Работа с реферативными сборниками, бюллетенями, проспектами. Работа с периодической печатью. Работа с книгой, монографией.	Доклад Опрос в устной форме Учебная задача	Опрос Вопрос 4

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 7	Первый экземпляр _____	КОПИЯ № _____

2	ПК-4 Для достижения ПК-4.1. знать: методологические принципы поиска, хранения, обработки и анализа информации	Основные формы представление информации. Естественные и искусственные языки. Текст как объект семантической обработки. Виды семантической обработки документов.	Доклад Опрос в устной форме Учебная задача	Опрос Вопрос 6
	Для достижения ПК-4.1. уметь: осуществлять поиск, хранение, обработку и анализ информации при проведении научных исследований в области лингвистики, межкультурной коммуникации и переводоведения	Формы, методы и средства идентификации, поиска и хранения информации. Особенности представления текстовой информации. Особенности представления числовой информации. Особенности представления графической информации. Особенности представления звуковой информации. Особенности представления мультимедийной информации.	Доклад Опрос в устной форме Учебная задача	Опрос Вопрос 3, 5
	Для достижения ПК-4.1. владеть: навыками практического использования современных информационных систем и баз данных при проведении научных исследований в области лингвистики, межкультурной коммуникации и переводоведения	Непрерывная и дискретная формы представления информации. Технологии предоставления информации. Технологии научно-исследовательского процесса. Электронные лингвистические ресурсы. Цифровые библиотеки. Информационные системы в филологических задачах.	Доклад Опрос в устной форме Учебная задача	Опрос Вопрос 1, 4

Типовые задания, критерии и показатели оценивания в рамках текущего контроля представлены в рабочей программе дисциплины (модуля). Полные комплекты оценочных средств и контрольно-измерительных материалов хранятся на кафедре.

3.2 Содержание оценочных средств

3.2.1. Опрос в устной форме

Указания: Дайте развернутый ответ на вопрос. Материал должен быть изложен последовательно, продемонстрирована высокая степень проработанности учебной, научной литературы, должны присутствовать выводы и примеры. Достаточный по объему (**5 минут звучания**). Ответ должен быть логически верно организован, четко структурирован, изложен с использованием терминов и понятий. Время подготовки **10 минут**.



МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение
ФГБОУ ВО «ЧелГУ»

Версия документа - 1

стр. 8

Первый экземпляр _____

КОПИЯ № _____

1. Информационные технологии. Основные характеристики. Виды. История и перспективы развития.
2. Информация и информационные системы. Основные характеристики. Виды. История и перспективы развития.
3. Современные подходы к анализу данных. Теоретическая и прикладная лингвистика.
4. Лингвистика в задачах информационного поиска. Информационно-поисковые языки как искусственные языки.
5. Современные технологии хранения данных. Виды словарей.
6. Автоматическая обработка текста. Текст и гипертекст.
7. Контент-анализ.

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 9	Первый экземпляр _____	КОПИЯ № _____

Ключи

1. Информационные технологии. Основные характеристики. Виды. История и перспективы развития.

Информационная технология (информационные технологии) - совокупность методов, производственных и программно-технологических средств, объединенных в технологическую цепочку, обеспечивающую сбор, хранение, обработку, вывод и распространение информации. Информационные технологии предназначены для снижения трудоемкости процессов использования информационных ресурсов.

В развитии информационной технологии можно выделить этапы. Каждый этап характеризуется определенным признаком. • Начальный этап развития ИТ (1950-1960-е годы) характеризуется тем, что в основе взаимодействия человека и ЭВМ лежат машинные языки. ЭВМ доступна только профессионалам • Следующий этап (1960-1970-е годы) характеризуется созданием операционных систем. Ведется обработка нескольких заданий, формулируемых разными пользователями; основная цель - наибольшая загрузка машинных ресурсов. • Третий этап (1970-1980-е годы) характеризуется изменением критерия эффективности обработки данных, основными стали человеческие ресурсы по разработке и сопровождению программного обеспечения. К этому этапу относятся распространение мини-ЭВМ. Осуществляется интерактивный режим взаимодействия нескольких пользователей • Четвертый этап (1980-1990-е годы) новый качественный скачок технологии разработки программного обеспечения. Центр тяжести технологических решений переносятся на создания средств взаимодействия пользователей с ЭВМ при создании программного продукта. Ключевое звено новой информационной технологии - представление и обработка знаний. Создаются базы знаний, экспертные системы. Тотальное распространение персональных ЭВМ.

Основные черты современных ИТ: • компьютерная обработка информации по заданным алгоритмам; • хранение больших объёмов информации на машинных носителях; • передача информации на значительные расстояния в кратчайшие сроки. Современная компьютерная технология базируется на трех основных принципах: • интерактивность - диалоговый режим работы с компьютером; • интегрированность с другими программными продуктами; • гибкость - возможность изменения, как данных,

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 10	Первый экземпляр _____	КОПИЯ № _____

так и постановок задач. Основу современных информационных технологий составляют три технических достижения: • появление новой среды накопления информации - магнитные и оптические диски; • развитие современных средств связи, в том числе и спутниковых; • постоянное совершенствование компьютеров и программного обеспечения для автоматизированной обработки.

Классификация информационных технологий. Основная классификация информационных технологий, это классификация, по технологии обработки: • Технологии обработки текстовой информации (текстовые редакторы, текстовые процессоры). • Технологии обработки числовой информации (табличные процессоры). • Технологии обработки графической информации (Графические редакторы, программы обработки векторной графики). • Технологии создания и обработки базы данных (системы управления базами данных).

Все базовые информационные технологии делятся на три группы: • Информационные системы. • Офисные технологии. • Телекоммуникации.

Требования к информационным технологиям: • малая стоимость, находящаяся в пределах доступности для индивидуального покупателя; • автономность в эксплуатации без специальных требований к условиям окружающей среды; • гибкость архитектуры, обеспечивающая ее адаптивность к разнообразным сферам применения: в управлении, науке, образовании, в быту; • "дружественность" операционной системы и прочего программного обеспечения, обуславливающая работу с ней пользователя без специальной профессиональной подготовки; • высокая надежность работы (более 8000 часов наработки на отказ).

Тенденции развития информационных технологий. В результате быстро набирающей темпы информационной революции одновременно происходит два процесса – резкое снижение цен на товары и услуги, связанные с современными технологиями, и стремительное, практически не имеющие прецедентов распространение ИТ в производственных системах и в сфере домашнего быта. Современные тенденции развития информационных технологий следует рассматривать как основной инструмент высоких технологий. Признавая поразительные технологические достижения наступившей эры информационных технологий, многие специалисты спрогнозировали дальнейшие тенденции развития их рынка. Это: 1) возрастание роли информационного



продукта; 2) развитие способности к взаимодействию (совместимости); 3) ликвидация промежуточных звеньев (непосредственность); 4) глобализация; 5) конвергенция. Рассмотрим эти тенденции. Возрастание роли информационного продукта. Информационный продукт - документированная информация, подготовленная в соответствии с потребностями пользователей и представленная в форме товара. Информационными продуктами являются программные продукты, базы и банки данных и другая информация. С одной стороны, ИТ расширяет кругозор людей, позволяет более эффективно использовать ресурсы, а с другой выполняет развлекательную функцию – обеспечивает досуг (Например, программное обеспечение и аппаратура для игры в боулинг). Способность информационных технологий к взаимодействию. Один из важнейших факторов для обеспечения совместимости взаимодействия — появление новых стандартов на программные и аппаратные средства, дисплеи, базы данных и сети. Всего несколько ключевых компонентов — микропроцессоры, локальные сети, робототехника, специализированные АРМ (автоматизированные рабочие места), датчики, программируемые контроллеры – превратили в реальность концепцию автоматизированного предприятия. Стандартизация зачастую связана с захватом большой рыночной доли какой-либо компанией, в результате чего её продукт становится стандартом для всех остальных. В качестве примера можно привести персональные компьютеры фирмы IBM, операционные системы корпорации Microsoft, локальные сети Novell, стандарты бытовой видеозаписи VHS фирмы JVC, Video-8 фирмы Sony.

****Способность к взаимодействию как концепция столь же привлекательна, сколь трудна в реализации. Трудности связаны одновременно с проблемами технологии и конкуренции. Ликвидация промежуточных звеньев. Это означает устранение тех стадий, Например, в банковской сфере уменьшается роль мелких банков, так как на внедрение информационных технологий требуются значительные ресурсы. Информационные технологии уже в настоящее время обеспечивают возможности для ликвидации промежуточных функций внутри компаний и между ними. **Телемаркетинг и система заказов "компьютер - компьютер" устраняют, например, промежуточные торговые организации. Покупатели, имеющие доступ к терминалам общего пользования (так называемым "электронным терминалам"), заказывают имеющиеся в продаже товары. Американские специалисты утверждают следующее: "Вводя новое оружие конкуренции**



в различные сферы деятельности, ИТ вызывают острую борьбу между фирмами". Intel предсказывает, что в конце концов "электронная" коммерция вытеснит среднее звено бизнеса (дистрибьюторов), которые сегодня играют важную роль в доведении всех видов продукции до конечного пользователя. Глобализация. Теоретически любой человек (или фирма) является сегодня возможным потребителем информации. На данный момент существует довольно жесткая конкуренция между основными производителями в сфере ИТ. К традиционно сильным производителям, таким как США, Япония, Франция, Великобритания и ФРГ, в последние годы добавились фирмы Австралии, Южной Кореи, Тайваня, Сингапура. Одной из главных причин интенсификации мировой конкуренции является распространение спроса на конкретные виды ИТ в мировом масштабе. Несмотря на различие рынков, продукция, пользующаяся спросом в Америке, фактически аналогична той продукции, на которую существует спрос в Японии и Европе. Сегодня в информационном бизнесе, как и в целом в сфере экономики, сложились три крупнейших центра — США, Япония, Западная Европа, между которыми и разворачивается вся конкурентная борьба. Конвергенция. Глобализация непосредственно связана с конвергенцией. Ранее сферу производства и сферу услуг можно было легко определить и дифференцировать. Однако сейчас традиционные представления меняются. Примерами являются системы кодировки брокерских операций, банковские аппараты, а также системы энергоконтроля. Кроме того, определенные виды продукции и услуг, выполняя одинаковые функции, становятся по существу взаимозаменяемыми. Конвергенция формирует сегодня свой собственный рынок ИТ, основными сегментами которого являются следующие элементы. 1) Потребительский сегмент включает передачу информации и развлечений и потребление их частными лицами. Такое потребление может осуществляться дома, в личном автомобиле, местном торговом центре или номере гостиницы. Многие индивидуальные услуги могут быть получены также служащими фирмы со своих АРМ. 2) Обеспечение бизнеса включает потребление продукции и услуг ИТ в ходе реализации различных видов деловой деятельности: закупки, производство-обслуживание, маркетинг, физическое распространение. Использование ИТ в этой области, как правило, чрезвычайно интенсивно. 3) Интеллектуальная работа относится к потреблению и передаче информации среди научных работников, менеджеров и других специалистов. Создание новой продукции и



необходимость быстрой ее конвергенции вызывают стремление к всестороннему сотрудничеству фирм в области ИТ. Конвергенция услуг является наглядным примером доминирующего значения информационного продукта, доступ потребителя к которому осуществляется посредством систем, способных к взаимодействию, что приводит, в свою очередь, к ликвидации промежуточных звеньев и традиционных каналов распространения, способствует глобализации. Проблемы внедрения новых информационных технологий Большинство решений о целесообразности внедрения той или иной ИТ принимаются в связи с жесткой необходимостью в оперативной и качественной управленческой информации или по причине усиления конкурентной борьбы на рынке.

Внедрение новых ИТ имеет ряд проблем, имеющих организационный и психологический аспект. К этим проблемам можно отнести: 1. временные и финансовые затраты 2. низкий уровень информационной культуры. Информационная культура — умение персонала целенаправленно работать с современными ИТ и использовать их для получения, обработки и передачи информации. Это характеристика информационной сферы жизнедеятельности людей. Информационная культура является продуктом разнообразных способностей человека, проявляющаяся в: • конкретных навыках по использованию технических устройств; • способности использовать в своей деятельности новые ИТ, базовой составляющей которых являются многочисленные программные продукты; • умении извлекать информацию из различных источников, представлять ее в понятном виде и уметь ее эффективно использовать; • владении основами аналитической переработки информации; • умении работать с различной информацией; • знании особенностей информационных потоков в своей области деятельности. 3. психологические проблемы, связанные с боязнью потери работы и снижением заработной платы как в случае удачного, так и в случае неудачного внедрения новых технологий.

2. Информация и информационные системы. Основные характеристики.

Виды. История и перспективы развития.

Информация (от лат. informatio, разъяснение, изложение, осведомленность) — по законодательству РФ - сведения о лицах, предметах, фактах, событиях, явлениях и

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 14	Первый экземпляр _____	КОПИЯ № _____

процессах независимо от формы их представления. С точки зрения информатики, информация обладает рядом фундаментальных свойств: новизна, актуальность, достоверность, объективность, полнота, ценность и др.

Классификация информации по форме представления • Текстовая • Числовая • Графическая • Звуковая • Видеоинформация

Под системой понимается совокупность связанных между собой и с внешней средой элементов или частей, функционирование которых направлено на получение конкретного полезного результата. Информационная система есть совокупность технического, программного и организационного обеспечения, а также персонала, предназначенная для того, чтобы своевременно обеспечивать надлежащих людей надлежащей информацией. Таким образом, основной задачей ИС является удовлетворение конкретных информационных потребностей людей в рамках конкретной предметной области. В соответствии с этим определением практически каждый экономический объект можно рассматривать как систему, стремящуюся в своем функционировании к достижению определенной цели. В качестве примера можно назвать систему образования, транспортную, экономическую и другие системы. Все информационные системы обладают рядом свойств, которые являются для них общими:

- ИС предназначены для сбора, хранения и обработки информации. Таким образом, в основе любой информационной системы лежат средства хранения и доступа к данным;
- ИС предназначены для конечного пользователя, не являющегося специалистом в области вычислительной техники. Из этого следует, что ИС должны включать в себя клиентские приложения, обеспечивающие интуитивно понятный интерфейс.

Классификация информационных систем. Существует множество способов классификации информационных систем. По классам реализуемых технологических операций информационные системы рассматриваются в программном аспекте и включают: текстовые редакторы, табличные процессоры, системы управления базами данных, графической редакторы, мультимедиа и другие системы. По характеру обработки данных ИС делятся на:

- информационно-справочные, или информационно-поисковые ИС, в которых нет сложных алгоритмов обработки данных, а целью системы является поиск и выдача информации в удобном виде;
- ИС обработки данных, или решающие ИС, в которых данные подвергаются обработке по сложным алгоритмам. К

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 15	Первый экземпляр _____	КОПИЯ № _____

таким системам в первую очередь относят автоматизированные системы управления и системы поддержки принятия решений. Управляющие системы обеспечивают получение решения, на основе которого человек принимает решение. Они предназначены для автоматизированного решения широкого круга задач управления. Примером могут служить система оперативного планирования выпуска продукции, система бухгалтерского учета. Классификация по сфере применения. Поскольку ИС создаются для удовлетворения информационных потребностей в рамках конкретной предметной области, то каждой предметной области (сфере применения) соответствует свой тип ИС. В качестве примера следующие типы ИС: • Экономическая информационная система — информационная система, предназначенная для выполнения функций управления на предприятии. • Медицинская информационная система — информационная система, предназначенная для использования в лечебном или лечебно-профилактическом учреждении. • Географическая информационная система — информационная система, обеспечивающая сбор, хранение, обработку, доступ, отображение и распространение пространственно-координированных данных (пространственных данных). Классификация по охвату задач (масштабности): • Персональная ИС предназначена для решения некоторого круга задач одного человека. • Групповая ИС ориентирована на коллективное использование информации членами рабочей группы или подразделения. • Корпоративная ИС в идеале охватывает все информационные процессы целого предприятия, достигая их полной согласованности, безызбыточности и прозрачности. Такие системы иногда называют системами комплексной автоматизации предприятия. Классификация по выполняемым функциям • производственные (конструкторские, технологические, управления производственным процессом и т.п.), • маркетинговые (анализа рынка, рекламные, снабженческие и т.п.), • финансовые (бухгалтерские, статистические, и т.п.) • кадровые системы. По сфере деятельности выделяют государственные, территориальные (региональные), отраслевые системы, системы объединений, предприятий или учреждений и технологических процессов. По масштабам применения выделяют настольные, офисные и корпоративные. • Настольные, в которых все компоненты находятся на одном компьютере; • Офисные (групповые) информационные системы ориентированы на коллективное использование информации членами рабочей группы (одного подразделения), строятся с использованием локальной



вычислительной сети. Такие системы позволяют повысить производительность труда секретарей и конторских работников и дают им возможность справляться с возрастающим объемом работ.

- Корпоративные информационные системы предназначены для управления территориально разделенной организацией. Такие системы строятся на основе Интернет–технологий. По вариантам использования сети в организации выделяют Intranet и Extranet. Intranet представляет собой внутренний корпоративный web-портал, который решает корпоративные задачи одной компании. Intranet-сайт доступен только в рамках локальной сети компании, включая удаленные филиалы. Extranet это защищенная от несанкционированного доступа корпоративная сеть компании, использующая интернет-технологии; это технология работы с сайтом компании внешних пользователей через сеть Интернет. Экстрасеть по сути является средой, связующей производство с поставщиками и потребителями. В зависимости от степени автоматизации информационных процессов бывают ручные, автоматические и автоматизированные системы. Ручные системы характеризуются выполнением всех операций человеком. Автоматизированные системы предполагают участие в процессе обработки информации и человека, и технических средств, причем главная роль отводится компьютеру. Автоматические информационные системы выполняют все операции по переработке информации без участия человека. По концепции построения бывают файловые системы, автоматизированные банки данных, банки знаний и хранилища данных.
- Файловые системы обычно обеспечивают хранение слабо структурированной информации, оставляя дальнейшую структуризацию прикладным программам. В таких системах сложно решить проблемы согласования данных в разных файлах, коллективного доступа к данным, модификации структуры данных.
- Банк данных – это совокупность баз данных, а также программные и другие средства, предназначенные для централизованного накопления данных и их дальнейшего использования на ПК. База данных – это совокупность самостоятельных материалов (статьи, расчеты, акты, судебные решения), систематизированные таким образом, чтобы они могли быть найдены и обработаны на ПК. В отличие от файловых систем, структура банков данных меньше зависит от прикладных программ, а все функции по работе с банками данных сосредоточены в системе управления базами данных, которая обеспечивает связь прикладных программ и пользователей.
- В банках знаний

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 17	Первый экземпляр _____	КОПИЯ № _____

информация о предметной области условно делится на две части: данные и правила. Данные хранятся в банке данных и, по сути, являются количественными и качественными характеристиками конкретных объектов. Правила – это сведения о закономерностях, представленные в виде программного обеспечения, позволяющие выводить новые факты из имеющихся. В отличие от данных, знания активны: на их основе формируются цели и выбираются способы их достижения. • Хранилища данных – это предметно-ориентированная информационная корпоративная база данных, специально разработанная и предназначенная для подготовки отчетов, анализа бизнес-процессов с целью поддержки принятия решений в организации. Они представляют собой автономный банк данных, в котором база данных разделена на два компонента. Оперативная база данных хранит текущую информацию. Квазипостоянная - содержит исторические данные, например, систематизированные годовые отчеты и балансы за все время существования предприятия. Подсистема оперативного анализа данных позволяет эффективно и быстро анализировать текущую информацию. Подсистема принятия решений пользуется обобщенной и исторической информацией, применяет методы логического вывода. Для общения с пользователем служит универсальный интерфейс.

Принципы создания информационных систем – ИС. Еще в 60-е годы прошлого столетия были сформулированы шесть основополагающих принципов, на которые необходимо опираться в процессе создания ИС. Развитие технической основы создания компьютеров и ИТ привело к переформулированию этих принципов и в ГОСТ РД 50-680-88 к ним отнесены следующие: системность, развитие (открытость), совместимость, стандартизация (унификация) и эффективность. 1) Принцип системности заключается в том, что должны быть установлены такие связи между структурными компонентами системы, которые обеспечивают цельность корпоративной системы и ее взаимодействие с другими системами. Нельзя разрабатывать какую-либо задачу автономно от других и реализовывать только отдельные ее аспекты. Задача должна рассматриваться комплексно со всеми возможными информационными связями. Пример: Отбор персонала на вакантные рабочие места. Ее решение должно осуществляться с учетом следующих моментов: ❖ использования результатов периодически проводимого профессионального и психофизиологического тестирования работников; ❖ анализа результатов периодически проводимой аттестации рабочих мест; ❖ анализа показателей трудовой

© ФГБОУ ВО «ЧелГУ»



дисциплины персонала; ❖ разработки общих и дополнительных критериев отбора (при наличии нескольких претендентов на одно рабочее место); ❖ использования банка данных претендентов, сформированного ранее; ❖ индивидуального собеседования; ❖ анализа анкетных данных и резюме (если претендент не является членом трудового коллектива). 2) Принцип развития (открытости) заключается в том, что внесение изменений в систему, обусловленных самыми различными причинами (внедрением новых информационных технологии, изменением законодательства, организационной перестройкой внутри фирмы и т. п.), должно осуществляться только путем дополнения системы без переделки уже созданного, т. е. не нарушать ее функционирования. Реализовать данный принцип на практике достаточно сложно. Необходимо разделить решаемые задачи на определенные группы и для каждой из них предусмотреть возможные направления развития (например, выход в глобальные сети, применение средств для сканирования документов, шифрование информации). 3) Принцип современности заключается в том, что при создании системы должны быть реализованы информационные интерфейсы, благодаря которым она может взаимодействовать с другими системами. В современных условиях это требует: ❖ дополнительных ужесточенных мер по защите информации; ❖ знания и соблюдения различного рода протоколов, регламентирующих все пины информационных обменов: ❖ знание сетевого этикета, предусматривающего такие правила, как: ❖ регулярная проверка своей электронной почты; ❖ периодическая чистка своего почтового ящика; ❖ корректность в составлении сообщений; ❖ указание координат для обратной связи и т.п. 4) Принцип стандартизации (унификации). При создании системы должны быть рационально использованы типовые, унифицированные и стандартизованные элементы, проектные решения, пакеты прикладных программ, комплексы, компоненты. Задачи необходимо разрабатывать таким образом, чтобы они подходили к возможно более широкому кругу объектов. 5) Принцип эффективности предусматривает достижение рационального соотношения между затратами на создание системы и целевыми эффектами, включая конечные результаты, отражающиеся на прибыльности и получаемые по окончании внедрения автоматизации в управленческие процессы. Перечень рассмотренных принципов создания корпоративных систем взят из ГОСТ. Однако к их числу с полным



правом можно отнести еще один из тех, которые были сформулированы в 60-е годы и по сей день не потеряли своей актуальности. Это принцип первого руководителя. Чрезвычайно важный принцип, распространяющийся на все сферы управленческой деятельности. Уровень компетентности руководителя любого уровня в производственных, административных, психологических и других вопросах определяет общие тенденции развития фирмы или ее подразделения и социально-психологический климат в коллективе. Известно, что устойчивое бесконфликтное взаимопонимание среди персонала способствует росту творческих начал и эффективной повседневной деятельности. И именно руководитель и первую очередь должен обеспечивать все элементы стабильности. Сформировать такой коллектив достаточно сложно и далеко не каждый руководитель способен это сделать. Напротив, негативное отношение руководителя к каким-либо нововведениям является тормозом в развитии творческой и профессиональной инициативы работников всех категорий.

Основные преимущества применения информационных систем:

- ИС автоматизирует применение математических методов к решению управленческих задач;
- ИС по крайней мере частично освобождает сотрудников от рутинного труда;
- ИС минимизирует вероятность появления ошибки в ходе передачи либо обработки информации;
- ИС снижает объем документов на бумаге;
- ИС совершенствует документооборот;
- ИС снижает затраты на производство товаров и услуг.

Внедрение информационных систем может способствовать:

- получению более рациональных вариантов решения управленческих задач за счет внедрения математических методов и интеллектуальных систем и т.д.;
- освобождению работников от рутинной работы за счет ее автоматизации;
- обеспечению достоверности информации;
- замене бумажных носителей данных на магнитные диски или ленты, что приводит к более рациональной организации переработки информации на компьютере и снижению объемов документов на бумаге;
- совершенствованию структуры потоков информации и системы документооборота в фирме;
- уменьшению затрат на производство продуктов и услуг;
- предоставлению потребителям уникальных услуг;
- отысканию новых рыночных ниш;
- привязке к фирме покупателей и поставщиков за счет предоставления им разных скидок и услуг.



3. Современные подходы к анализу данных. Теоретическая и прикладная лингвистика.

Большой энциклопедический словарь «Языкознание» дает следующее определение лингвистической науке: «Лингвистика (языкознание, языковедение) – это наука о естественном человеческом языке вообще и обо всех языках мира как индивидуальных его представителях». Лингвистика в широком смысле занимается познанием языка и передачей результатов этого познания другим людям для их практических целей. Интересен факт, что сам термин «лингвистика» (linguistics) вошел в научный обиход в 1847 году, хотя термин «linguist» (в значении «a student of language») возник на 200 лет раньше. Языковедение возникло в связи с практическими потребностями людей и тесно связано с появлением письменности в период, по разным оценкам, от 2 до 5 тысяч лет до нашей эры на Древнем Востоке. Одними из первых лингвистических продуктов ученые считают шумерские глоссы (25 в. до н. э.), а первым серьезным научным трудом – формальные грамматики санскрита древнеиндийского ученого Панини (V в. до н. э.). Само же языковедение стало оформляться как отдельная наука лишь в XVII–XVIII веках, что было обусловлено ее огромной сложностью и недостатком знаний о языке – объекте этой науки. В XIX–XX вв. отмечается подъем в области языкознания, появляются известные лингвистические школы (американская, пражская, 8 копенгагенская, московская и др.) и направления. Огромный вклад в лингвистическую теорию вносит Фердинанд Де Соссюр, который разрабатывает основы структурализма, или структурную лингвистику, развивает теорию языка и речи [Шарафутдинова, 2012, с. 249-255]. Практическое направление, называемое сегодня «прикладная лингвистика», активно стало развиваться в середине XX века в связи с появлением компьютерных технологий и усовершенствованием технических средств. История лингвистики как науки, различные лингвистические школы, этапы развития лингвистических учений рассматриваются подробно в рамках курса «История лингвистических учений» [Шарафутдинова, 2012, часть 2]. Языкознание изучает различные стороны языка, рассматривая их в рамках своих отдельных научных направлений. Например: Фонетика и фонология – научные направления языкознания, изучающие звуковую систему языка и особенности ее функционирования. Лексикология изучает словарный состав языка и закономерности его развития. Грамматика, как наука о



способах и средствах построения и изменения слова, о способах и средствах построения предложения, подразделяется на две области – морфологию и синтаксис. Морфология – учение о способах и средствах построения и изменения слов. Синтаксис – учение о способах и средствах построения предложений. Семантика – наука о смысле (значениях) слов и предложений, частей речи и членов предложений. Прагматика – учение об условиях и целях коммуникации, влияющих на понимание, так называемое «изучение языка в контексте», изучение отношений между средствами языка и теми, кто этими средствами пользуется. Дискурс изучает и анализирует характерные для данного вида текстов (или коммуникантов) особенности лексики, синтаксиса, семантики либо прагматики языковых единиц, проявляющиеся в актуальных коммуникативных актах, речи и письменных текстах (например, спортивный дискурс). Мы также будем рассматривать лингвистику как сложную и разветвленную систему дисциплин как теоретического, так и прикладного характера. Приведем лишь некоторые из таких дисциплин для демонстрации разнообразия языковедческих направлений. Примеры лингвистических дисциплин: 1. Общее языкознание занимается проблемами, касающимися всех языков, исследует сущность и природу языка вообще, проблему его происхождения и общие законы его функционирования, структуру, классификацию языков. Общее языкознание разрабатывает методы исследования языков, формулирует языковые универсалии, т.е. положения, действительные для всех или большинства языков мира. 2. Частное языкознание изучает один или группу родственных языков (например, германистика, романистика, тюркология, балканистика и т.д.). В зависимости от целей и задач исследования частное языкознание может быть синхроническим, описывающим фактор языка в какой-либо момент истории, в одной временной плоскости, чаще всего современное его состояние, и диахроническим, прослеживающим историческое развитие языка. 3. Сравнительно-историческое языкознание (компаративистика) занимается установлением степени родства между языками (построением генеалогической классификации языков), реконструкцией первых языков, или праязыков, исследованием различных процессов в истории языков, их групп и семей, этимологией слов. 4. Типология (универсализм) занимается выяснением наиболее общих закономерностей различных языков, не связанных между собой общим происхождением или взаимовлиянием. В случае, если некоторое явление выявляется в представительной



группе языков, оно может считаться типологической закономерностью, применимой к языку как таковому. Типологический анализ возможен на уровне звука (фонетическая типология), на уровне слова (морфологическая типология), предложения (синтаксическая типология) и далее (типология текста или дискурса). 5. Диалектология (ее вариант – лингвистическая география) – изучает местные территориальные разновидности одного языка. 6. Когнитивная лингвистика – междисциплинарное направление языкознания, тесно связанное с семантикой, которое исследует проблемы соотношения языка, сознания и мышления, роль языка в познании, понимании и отражении окружающей действительности. 7. Паралингвистика – изучает неязыковые (невербальные) средства в речи, передающие совместно с вербальными смысловую информацию в составе речевого сообщения, а также совокупность таких средств. Изучение невербального общения является важным компонентом межкультурной коммуникации. Паралингвистику также считают разделом невербальной семиотики. 8. Психолингвистика – связана с изучением речи, речеобразования (чаще всего конкретного индивида) для решения тех проблем, которые не могут быть решены ни в психологии, ни в лингвистике. Психолингвистика – самостоятельное научное направление, возникшее на стыке лингвистики и психологии. Объединение этих двух дисциплин в одну пограничную дисциплину позволяет использовать понятийный аппарат лингвистики для описания языковой формы речевых высказываний и понятийный аппарат психологии для описания и объяснения психических процессов производства и восприятия речи. В рамках психолингвистики изучается, например, детская речь, речевые отклонения при психических заболеваниях. 9. Социолингвистика – изучает роль языка в обществе, воздействие общества на язык, язык в связи с социальными условиями его существования, комплексом внешних обстоятельств, в которых реально функционирует и развивается язык: общество людей, использующих данный язык, социальная структура этого общества, различия между носителями языка в возрасте, социальном статусе, уровне культуры и образования, месте проживания, а также различия в их речевом поведении в зависимости от ситуации общения. В рамках социолингвистики активно развивается, например, политическая лингвистика. 10. Прикладная лингвистика – к этой области относятся практические направления языкознания, часто междисциплинарного характера, например, компьютерная лингвистика (машинный перевод, лингвистические



основы информатики, информационный поиск и др.), лингводидактика (обучение языку), дешифровка (исследование текстов на незнакомом коде, или языке, для получения информации), математическая лингвистика (например, квантитативная лингвистика, проводящая статистический анализ языка и текстов, занимающаяся построением и использованием структурно-вероятностных моделей языков).

Теоретическая и прикладная лингвистика Лингвистику условно принято подразделять на теоретическую, иногда ее еще называют «научная лингвистика», или «теория языкознания» (в рамках этого направления рассматриваются различные научные концепции, лингвистические теории, лингвистические школы, язык с точки зрения его структуры и систем), и прикладную (практическую) лингвистику. Одним из важнейших направлений развития лингвистики является прикладная лингвистика, одной из главных задач которой является решение и оптимизация традиционных лингвистических задач, как, например, перевод или обучение языкам. Обеспечению практических потребностей людей служат такие разделы прикладной лингвистики, как написание практических грамматик, практическая фонетика, практическая лексикография (создание учебных или отраслевых словарей). Каждое из направлений практической лингвистики обычно имеет свое отражение в сфере теоретической лингвистики. Так, например, проблемы перевода изучаются в рамках такой дисциплины, как переводоведение (Translation Studies), которая формирует свою теорию перевода, различные переводческие концепции. В российской традиции изначально было принято широкое понимание термина «прикладная лингвистика», на что указывает профессор А. Н. Баранов в своем вводном курсе и пособии «Введение в прикладную лингвистику» [Баранов, 2003]. И в такой широкой интерпретации прикладная лингвистика является комплексной научной дисциплиной, применяющей лингвистические знания в различных ситуациях для решения разного рода практических задач (таких, например, как машинный перевод, техническая коммуникация, распознавание и синтез речи, информационный поиск и др.). Следует также отметить, что за рубежом термин «Applied Linguistics» иногда трактуется узко, а под ним часто понимается только сфера преподавания иностранных языков, т. е. сфера подготовки преподавателей иностранных языков, изучение методики лингводидактики. По мнению многих ученых, настоящий период развития языкознания характеризуется серьезным влиянием на лингвистическую теорию прикладной

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 24	Первый экземпляр _____	КОПИЯ № _____

лингвистики. По словам известного ученого В. А. Звегинцева, прикладная лингвистика является «полигоном, где проходят испытания как частные гипотезы, так и глобальные теоретические построения...» [Звегинцев, 2008]. Также, когда говорят о проверке лингвистических теорий на практике, то имеют в виду не только, например, создание практических грамматик и словарей, но и их эффективное приложение (использование) в обучении и преподавании. Современная прикладная лингвистика активно развивается и решает многие проблемы обработки устной и письменной речи, извлечения и использования информации из текстов, повышает эффективность коммуникации, в том числе в сфере информационных и компьютерных технологий. Прикладная лингвистика представляет собой ту сферу, где реально проводятся лингвистические эксперименты, имеющие целью верификацию положений теории языкознания и проверку эффективности лингвистических продуктов, создаваемых разработчиками. Как говорит профессор СПбГУ А. С. Герд: «К счастью, прикладная лингвистика – наука экспериментальная» [Герд, 2005, стр. 5].

Задачи и направления прикладной лингвистики. Определим круг некоторых основных задач прикладной лингвистики: 1. Перевод с/на иностранный язык; 2. Обучение иностранному языку (методики обучения), лингводидактика; 3. Поддержка коммуникации с помощью технических средств; 4. Лингвистические основы информатики (создание искусственных языков, например, языков программирования, web-разработки); 5. Лингвистические основы информационного поиска; 6. Аннотирование, реферирование, классификация текстов; 7. Составление словарей (практическая лексикография); 8. Терминология и терминография (упорядочение, стандартизация и унификация научно-технической терминологии); 9. Распознавание символов текста; 10. Распознавание и синтез устной речи.

В последние 50 лет внедрения новых информационных технологий во все сферы человеческого общения прикладная лингвистика развивается по направлению автоматизации основных задач, оптимизации коммуникации, т. е. развивается такое направление прикладной лингвистики, которое называли компьютерной лингвистикой. Сюда входят, например: • Машинный перевод (Machine Translation); • Компьютерная лексикография (подготовка электронных словарей); • Компьютерная лингводидактика (CALL/T – Computer Assisted Language Learning and Teaching); • Автоматическая



обработка естественных языков (NLP – Natural Language Processing). В рамках прикладной лингвистики для решения вышеуказанных задач необходимо было выработать эффективные методы исследования языка. Среди таких классических методов можно отметить, например, вероятностно-статистические методы, логико-математические методы, методы моделирования, корпусного анализа, контент-анализа и многие другие, которые активно и плодотворно использовались в других науках и нашли свое эффективное приложение в лингвистике.

Лингвистика – сложная многоплановая область знаний и приложений, которая соприкасается как со многими гуманитарными науками, так и с естественнонаучными. В последние годы появились смежные лингвистические направления на стыке тех или иных наук, которые активно развиваются за счет новых теоретических и практических приложений. Укажем лишь некоторые: • Социология, Политология - > Социолингвистика, Политическая лингвистика; • Филология - > Лингвистика текста; • Психология - > Психолингвистика; • Логика - > Когнитивная лингвистика, Информационный поиск, Моделирование семантики и знаний; • Математика - > Математическая лингвистика, Лингвостатистика; • Биология - > Фонетика; • Физика - > Распознавание и синтез речи; • История, археология - > Этнолингвистика, Дешифровка, этнография Сравнительно-историческое языкознание; • Юридическая наука - > Юридическая лингвистика (рассматривает язык права, проводит лингвистические экспертизы текстов и высказываний); • Информатика - > Компьютерная лингвистика (например, ее раздел – лингвистические основы информатики, связанный с разработкой и обработкой искусственных языков программирования); • Семиотика - > Знаковая теория языка, WEB-дизайн.

4. Лингвистика в задачах информационного поиска. Информационно-поисковые языки как искусственные языки.

Теория и практика информационного поиска стала развиваться в середине XX века. К концу века с внедрением мировых информационных компьютерных сетей и их информационно-лингвистического обеспечения эта область стала отдельным признанным научно-практическим направлением компьютерной науки со своей теорией

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 26	Первый экземпляр _____	КОПИЯ № _____

информационного поиска и многочисленными практическими разработками, продуктами и технологиями. Поиск информации (Information Retrieval) – это процесс отыскания в некоторой системе хранения информации таких документов (текстов, записей и т. д.), которые соответствуют поступившему запросу. В качестве таких средств хранения и поиска информации выступают информационно-поисковые системы (IRS/Information Retrieval Systems), элементами которых являются структурированный массив документов (база данных/индекс), выступающих как объект поиска, разнообразные технические и программные средства, как например, программы-роботы, а также информационно-поисковый язык, задающий правила индексирования документов, правила поиска. По В. П. Захарову «Информационно-поисковая система (ИПС) – это упорядоченная совокупность документов (массивов документов) и информационных технологий, предназначенных для хранения и поиска информации – текстов (документов) или данных (фактов). Информационно-поисковыми системами являются любые определенным образом организованные хранилища информации. Причем информационно-поисковые системы могут быть и неавтоматизированными. Главное – это целевая функция: хранение и поиск информации». При вводе документа в базу данных ИПС его индексируют (поэтому саму базу поисковой системы часто называют индексом). Процесс индексирования в основном связан с определением и выборкой ключевых слов обрабатываемых документов и выражением их формально в виде поискового образа. Так база данных ИПС состоит из множества индексных поисковых образов. Непосредственно при поиске производится сопоставление вашего запроса, т. е. того, что вы указали в запросе с поисковым образом, т. е. тем, что хранится в индексе.

В зависимости от объекта хранения и типа запроса часто в учебной литературе различают два вида информационного поиска: • документальный (когда пользователь ищет какой-либо текст/документ); • фактографический (когда пользователь ищет какие-либо фактические данные, например, «День рождения Анджелины Джоли»). Характеристики информационного поиска – это такие его семантические показатели как: • полнота выдачи информации; • точность ее выдачи; • потери информации; • информационный шум. Полнотой поиска (Recall) называется мера, вычисляемая как отношение количества выданных релевантных документов к общему числу релевантных документов, содержащихся в базе информационно-поисковой системы. Точность поиска

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 27	Первый экземпляр _____	КОПИЯ № _____

(Precision) – это отношение количества выданных системой релевантных документов к общему числу документов в выдаче. Особенно интересно нам понятие «релевантность» как фундаментальное понятие теории информационного поиска. Согласно определению, документ, центральный предмет или тема которого в целом соответствует смысловому содержанию информационного запроса, называется релевантным, а свойство смысловой близости между документом и информационным запросом – релевантностью. По многим, чаще всего субъективным, причинам релевантность была и остается основной проблемой информационного поиска.

Теория и практика информационного поиска тесно связана с лингвистикой, т. к., во-первых, основной объем информации и тексты документов представлены на естественных языках, а для извлечения информации и индексирования текстов необходимы знания автоматической обработки языков, например, правила компьютерной морфологии. Во-вторых, информационные компьютерные системы построены и работают на базе искусственных языков. Из определения информационно-поисковой системы видно, что ее основным лингвистическим средством является специализированный искусственный язык. Информационно-поисковый язык (ИПЯ) – это специализированный искусственный язык, предназначенный для: 1) описания информационных запросов к информационно-поисковым системам (язык запросов), 2) описания формальных характеристик документов в виде поискового образа, хранящегося в базе системы (язык индексирования). Необходимость внедрения искусственных языков вызвана необходимостью устранения избыточности естественного языка для информационного поиска, а также ликвидации языковой синонимии, омонимии и неоднозначностей разного рода. Информационно-поисковый язык, как и любой язык, состоит из фиксированных единиц, например, имеет свой словарь и синтаксис, и является искусственным языком, т.е. ограниченным по своей форме и структуре стоящими перед ним задачами на поиск. Рассмотрим каждый из вариантов ИПЯ более подробно: языки запросов кажущиеся простыми на первый взгляд языки запросов представляют собой довольно комплексные системы правил и процедур. Как правило, модель языка запросов включает следующие элементы: • Поисковые единицы (ключевые слова, выражающие информационную потребность пользователя). • Средства лемматизации лингвистических единиц запроса (приведение слов к

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 28	Первый экземпляр _____	КОПИЯ № _____

нормальной словарной форме). • Специальные поисковые (булевские) операторы, типа OR, NOT, AND. • Средства линейной грамматики (операторы расстояния, позиционные операторы, например, title:). • Дополнительные условия поиска: ограничение области поиска по языку, региону, дате создания документа и т. п. Запрос на поиск чаще всего ограничен по языку и может иметь определенное формальное представление. Самый частотный способ поискового запроса – через ограниченный набор ключевых слов, задаваемый пользователем в поисковой строке. Например, найти эту книгу в Интернете можно, задав ключевые слова Соснина введение прикладная лингвистика в поисковых системах Яндекс, Google и т. п. Кроме стандартной практики в поисковых системах предусмотрена функция расширенного языка запросов, ее можно легко найти на сайтах поисковых машин – <http://yandex.ru/search/advanced> или http://www.google.ru/advanced_search. Расширенный поисковый язык включает множество операторов для улучшения качества и сужения зоны поиска релевантной информации. Например: • "чемпионат мира 2014" OR "олимпийские игры 2014"; • мумий тролль мультфильм -рок -лагутенко – Поиск в Яндексе или Google только мультфильма, а НЕ (-отрицание) группы Ильи Лагутенко; 50 • социолингвистика -site:ru.wikipedia.org (ищем информацию везде кроме сайта русской Википедии); • date:ГГГГ{*|ММ{*|ДД}} – Поиск в Яндексе только по страницам, дата которых удовлетворяет заданному условию (например, ищем труды профессора УлГТУ только за 2012 г. Шарафутдинова Н. С. date: 2012).

Языки индексирования. В литературе по теории информационного поиска обычно выделяют следующие распространенные виды информационно-поисковых языков для индексирования документов: 1. Языки классификаций – иерархические; алфавитно-предметные и др.; 2. Дескрипторные языки. Иерархические ИПЯ (иерархия – классификация от общего к частному). Такая организация используется для поиска книг в библиотеке (например, Универсальная десятичная классификация – УДК). Дерево классов Алфавитно-предметные ИПЯ представляются как алфавитный список ключевых слов какого-либо документа с пометами (например, алфавитно-предметный указатель в конце книг). Используется для построения различных указателей, каталогов, картотек.

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 29	Первый экземпляр _____	КОПИЯ № _____

Дескрипторные языки – наиболее естественная и популярная форма языков индексирования для выражения поисковых образов и в настоящее время широко используется в современных информационно-поисковых системах (в частности во многих поисковых системах сети Internet). В таких языках используется принцип координатного индексирования, т. е. перечисляются ключевые слова или дескрипторы, которые выражают центральную тему или целостную характеристику искомого объекта. При этом используется принцип логического умножения понятий, в результате которого из простых лексических единиц строятся более сложные, выражающие более узкие понятия. Этот принцип описания содержания документов через перечисление ключевых слов существует издавна. Например, пересечением понятий ПАРИЖ и ДОСТОПРИМЕЧАТЕЛЬНОСТИ, заданных в запросе, порождается новое более узкое понятие ДОСТОПРИМЕЧАТЕЛЬНОСТИ ПАРИЖА. В качестве лексических единиц в дескрипторных ИПЯ выступают дескрипторы – имена понятий или классов понятий, которые явно перечисляются в дескрипторном словаре. Это слова (или словосочетания), выбранные в качестве представителей классов и групп синонимичных слов и словосочетаний. Как правило, это существительные или номинативные выражения. Списки дескрипторов организованы в специальные семантические словари (поисковые тезаурусы). Поисковые тезаурусы (одноязычные или многоязычные) – специальные словари для информационного поиска по массивам естественно-языковых документов, организованные по принципу сопоставления слов с их понятиями. Их структура и разработка часто стандартизируются, см., например, ГОСТ 7.25-2001. Такой дескрипторный словарь используется как средство лексического контроля (например, снятия омонимии, синонимии единиц) при индексировании документов и запросов. Например, при индексировании все синонимы запроса и поискового образа представляются одной и той же лексической единицей – дескриптором (ср. лингвистика – языкознание языковедение, наука о языке).

Структурно поисковый тезаурус представляет собой алфавитный список дескрипторов вместе с их словарными статьями (гнездами). Словарная статья обычно содержит: • заглавный дескриптор; • ключевые слова или словосочетания, входящие в гнездо данного дескриптора (условные синонимы); • «вышестоящие» дескрипторы (находящиеся с данным в отношении «род–вид», «часть–целое»); • «нижестоящие»

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 30	Первый экземпляр _____	КОПИЯ № _____

дескрипторы (находящиеся с заглавным дескриптором в отношении «вид–род», «целое–часть»); • ассоциативные дескрипторы (связанные с данным другими разнообразными отношениями: причина–следствие, процесс–объект, свойство–носитель свойства, функциональное сходство). Указанные отношения приводятся с пометами, чаще всего так: ДЕСКРИПТОР с – ключевые слова (синонимы) в – родовые слова (дескриптор, подчиняющий данный) н – видовые слова (дескриптор, подчиненный данному) а – ассоциации (отношения).

5. Современные технологии хранения данных. Виды словарей.

Лексикография (от греч. *lexis* 'слово' и *grafia* 'писание, наука'), будучи одним из важных направлений прикладной лингвистики, занимается теорией и практикой составления словарей. В лексикографии выделяют два научно-практических направления: традиционная и машинная (компьютерная) лексикография. Традиционная лексикография имеет глубокие исторические корни и в большей мере занимается теорией и практикой составления «традиционных» словарей. Самые ранние глоссы (от греческого *glossa* 'язык, слово', словарные пометы о значениях незнакомых слов) известны с глубочайшей древности (например, шумерские глоссы XXV в. до н. э.). Поэтому, самыми первыми типами словарей многие ученые считают глоссарии, написанные от руки списки иностранных и необычных слов, с которыми приходилось сталкиваться в манускриптах на древних языках. Машинная лексикография (Computational Lexicography) занимается автоматизацией подготовки словарей и решает задачи разработки электронных словарей. В отличие от традиционной, машинная (компьютерная) лексикография – относительно молодая наука, реализующая традиционные наработки в технических средах и создающая разнообразные электронные словари. К основным задачам компьютерной лексикографии относятся также задачи разработки технологий составления электронных словарей и управления терминологией (Terminology Management).

В настоящее время можно выделить три основных направления компьютерной лексикографии: 1) автоматическое получение из текста различных словарей (например, терминологических, частотных словарей, словарей конкордансов и др.); 2) создание словарей, являющихся электронными версиями традиционных словарей (например,



словарь Даля), или комплексных электронных лингвистических словарей для традиционных словарных работ, например, известный словарь LINGVO [<http://lingvopro.abbyyonline.com/ru>]. Большой выбор такого рода открытых словарей доступен через поисковые системы, например, <http://slovari.yandex.ru/>. 3) разработка теоретических и практических аспектов составления специальных компьютерных словарей, например, для информационного поиска, машинного перевода (например, поисковые тезаурусы, словари стоп-слов, словари основ и флексий для морфологического анализа/синтеза). Таким образом, мы видим, что основным объектом интересов лексикографии является такой продукт как словарь и все задачи, связанные с разработкой разного типа словарей.

Словарь – определенным образом организованное собрание слов, обычно с приписанными им комментариями, в которых описываются особенности их структуры и/или функционирования. Помимо слов, объектами словарного описания могут выступать их компоненты (таковы, например, словари морфем), словосочетания различных типов, устойчивые группы – пословицы, поговорки, цитаты и т. п. В другом значении термин «словарь», или лексикон, обозначает всю совокупность слов некоторого языка (иначе говоря, его лексику) и противопоставляется термину «грамматика», обозначающему совокупность правил построения из слов более сложных языковых выражений. Таким образом, под словарем понимают:

- полный словарный состав языка;
- упорядоченное для решения практических задач множество лексических единиц;
- справочную книгу слов, расположенных в определенном порядке, дающую кому-либо информацию о том или ином слове.

Основная задача словаря – это представление либо описание лексики языка и ее особенностей для решения конкретных задач. Это сложнейшая проблема, так как лексика языка имеет тенденцию увеличиваться и качественно изменяться. Вот как метафорично говорит В. Селегей в статье «Электронные словари и компьютерная лексикография»: «Многие словари, основной корпус статей которых сформировался в языковой атмосфере середины века, представляют собой лексикографические музеи (а то и терминологические кладбища, если говорить о специализированных словарях)».

Классификация словарей. Словари нелингвистические (энциклопедии и т.п.), общие лингвистические словари (толковые, переводные, орфографические, частотные и

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 32	Первый экземпляр _____	КОПИЯ № _____

др.), специальные (синонимов, антонимов, словарь пословиц и поговорок, сленга, словари иностранных слов и др.).

Основной структурной единицей словаря (как книги) является словарная статья, организация и моделирование которой является зачастую сложнейшей прикладной проблемой и задачей лексикографии. Словник словаря – это перечень терминов словаря без их толкований. Важным вопросом при составлении словарей также является порядок расположения словарных статей, чаще всего это алфавитный порядок или предметный (тематический), при котором слова группируются по темам или графически (например, тематический визуальный словарь). Электронные словари часто представляют из себя сложный комплекс компьютерных программ – лингвистических платформ. Примером тому служит ABBYY Lingvo Content – профессиональное приложение, которое позволяет легко создавать новые словари и глоссарии, редактировать их и публиковать в бумажном и электронном виде на сайтах, на корпоративных порталах, а также в виде приложений для PC, Mac, смартфонов и мобильных устройств (http://www.abbyy.ru/lingvo_content).

Отметим некоторые особенности автоматических словарей: • кроме словарной базы данных (перечень слов по алфавиту) для автоматической словарной работы необходимы специальные алгоритмы и программы, например, морфологической нормализации, или лемматизации. Лемматизация – это приведение разных форм слова к его канонической (исходной) форме (одна из задач компьютерной морфологии); • в машинных словарях присутствуют не только перечни отдельных слов, но и до 50 % словосочетаний (особенно в терминологических словарях, которые очень важны для перевода специальных и технических текстов); • электронные словари получили в настоящее время большое распространение в силу их доступности для широкого круга пользователей через различные мобильные и персональные устройства. Примечание: следует отличать автоматические словари и системы машинного перевода (как ни странно, их часто путают обыватели). Последние лишь включают первые, являясь значительно сложнее.

Класс идеографических словарей (предметные, тематические), к которым часто относят и тезаурусы, – это особого рода словари, организованные, во-первых, по тематическому принципу, и, во-вторых, по принципу иерархии отношений и «от смысла

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 33	Первый экземпляр _____	КОПИЯ № _____

к слову», т. е. идеографические словари ориентированы на семантику языка, и каждый такой словарь – это некоторая семантическая модель лексики, построенная на иерархических отношениях типа «род – вид», «часть – целое», «синонимы» и т. п. Вот что говорится в известном издании про идеографические словари «Альтернативой алфавитному расположению слов является размещение их по смысловой близости. Словари, в которых лексика располагается на основании этого критерия, получили название идеографических (от греч. *idea* – понятие, идея, образ и *grapho* – пишу)». Тезаурус – это также идеографический словарь, но имеющий четкую сложную иерархию отношений. В словарных статьях тезауруса отражены существенные для данного термина связи с другими понятиями, иными словами – это маленькая энциклопедия. Самый известный классический словарь-тезаурус – это тезаурус П. Роже «*Thesaurus of English words and phrases*», опубликованный в 1852 г., неоднократно переиздававшийся как классический лексикографический труд и представленный в сети Интернет. П. Роже систематизировал лексику английского языка по категориальным группам, каждая из которых представлена именем понятия, которых сначала насчитывалось 1000. Слова расположены в алфавитном порядке, далее идут синонимы слов по частям речи (существительные, глаголы, прилагательные, наречия), антонимы и затем списки родственных слов. Многими отмечается, что ценность этого тезауруса в его естественности, в том, что это описание общей понятийной лексики языка, а также в том, что его можно привлекать к использованию в компьютерных системах как семантическое средство. Направление, связанное с разработкой семантических словарей, активно развивается. Существуют глобальные проекты семантических словарных баз типа WORDNET, а также такие любопытные версии словарей как визуальные тезаурусы, наглядно представляющие визуальные карты выбранных слов.

6. Автоматическая обработка текста. Текст и гипертекст.

С появлением компьютерных сетей и новых информационных технологий несколько трансформировались традиционные лингвистические понятия, в частности, понятие текста. Под текстом (лат. *textus* – связанность, материал, сплетение), письменным или устным, принято понимать логически связанную последовательность лингвистических знаков. Основные характеристики текста: связанность; осмысленность;



цельность (текст должен быть закончен по смыслу). Кроме того, классически текст имеет линейную структуру. С середины 80-х годов XX века, когда быстрыми темпами стали развиваться компьютерные и телекоммуникационные сети (WWW, Internet) и вместе с тем глобальная коммуникация, в научный обиход вошло новое понятие гипертекста (Hypertext). Оказалось, что для решения многих информационных задач нелинейный и мультимедийный характер текста может быть эффективнее. Гипертекст – это технология организации информации и особым образом структурированный текст, разбитый на отдельные блоки, имеющий нелинейное представление, для эффективной презентации информации в компьютерных средах. Для стандарта представления информации в сетях стали разрабатываться специализированные языки. Например, HTML – HyperText Markup Language – язык гипертекстовой разметки документов. При работе с текстом можно решать множество задач от общефилологических до статистических. Условно направление обработки текста с помощью доступных компьютерных и информационных технологий можно назвать автоматической обработкой текста, т. е. это любое преобразование текста на естественном или искусственном языках с помощью компьютера. Автоматическая обработка текстов осуществляется, как мы заметили раньше, в технологиях информационного поиска (при индексировании текстов), машинного перевода, словарной работе.

- Автоматизированные операции по печати, редактированию и верстке текста. Первоначально пользователю были доступны примитивные операции по обработке текста в простых компьютерных программах-редакторах (Lexicon), впоследствии требования к редактированию и представлению документов возрастали, что привело к созданию усовершенствованных систем типа WordProcessors (Microsoft Word и др.). Сейчас это развитые программы с функциями обработки таблиц, графики, проверки орфографии и стилистики. Кроме того, есть специальные издательские системы для профессиональной верстки документов, газет, рекламной продукции, книг (например, PageMaker или продвинутые пакеты типа Adobe® Digital Publishing Suite).

- Распознавание текста. Распознавание текста в этом смысле привязано к задаче распознавания символов (Character Recognition) и реализуется эффективнее для печатного текста, чем для рукописного. С работой лингвистического распознавателя пользователь сталкивается при сканировании текста, например, через платформу

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 35	Первый экземпляр _____	КОПИЯ № _____

ABBYY FineReader. В настоящее время задача распознавания печатного текста практически решена в силу ограниченного набора печатных гарнитур. Для рукописного текста (технологию называют ICR, т. е. Intelligent Character Recognition) распознавание графем гораздо сложнее, так как нужны более совершенные алгоритмы распознавания образов.

- Автоматическое реферирование и аннотирование текстов. Автоматическое реферирование (Automatic Text Summarization) – это составление коротких изложений материалов, аннотаций или дайджестов, т. е. извлечение наиболее важных сведений из одного или нескольких текстовых документов и генерация краткого содержания анализируемого объекта. Аннотирование, в отличие от реферирования, предполагает еще большее сжатие содержания исходного текста. Существует много способов решения этой задачи прикладной лингвистики. Как правило, автоматическое реферирование основано на выборке ключевых фрагментов документов – выделении наиболее информативных фраз и элементов – и их сборке. При автореферировании активно используется статистическая информация по информативности разных элементов текста по частоте их появления в нем, информативность элемента текста также зависит от его позиции в документе или текстовых маркеров (например, термины выделены курсивом или жирно), которые характеризуют их содержательную значимость. Существуют как коммерческие, так и открытые, например, Open Text Summarizer, отдельные продукты для реферирования, а также встроенные функции автореферата (например, в MS Word).

- Корпусная лингвистика. Корпусная лингвистика (Corpus Linguistics) – целое направление прикладной лингвистики, активно развивающееся с конца XX века, а в России – с начала XXI века, например, Национальный корпус русского языка, стал разрабатываться лишь в 2004 году. В настоящее время электронных корпусов текстов для разных языков и приложений существует большое множество. Задачи корпусной лингвистики связаны с разработкой технологий построения электронных лингвистических ресурсов особого типа – корпусов текстов (Corpora) – и решением задач разного рода на базе этих текстов. В основном, такие коллекции и массивы текстов отражают реальное функционирование того или иного языка, а их перенос в компьютерные среды активизировал их практическое и широкое использование в прикладной лингвистике. Корпусная лингвистика дает материал для различного рода

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 36	Первый экземпляр _____	КОПИЯ № _____

исследований языка и его вариантов и определяет основной метод анализа текстов и языка на базе корпусов. Одной из важных особенностей метода анализа на базе корпусов является исследование не только чисто лингвистических явлений (грамматических или лексических функций слов, их связей с другими лексемами), но и таких явлений, как, например, частотности лексем или грамматических конструкций в тех или иных жанрах, диалектах. Корпусный подход, или метод лингвистического исследования, основанный на корпусах текстов, ориентирован на прикладное изучение языка, его функционирования в реальных средах и текстах, что важно, например, для преподавания языка и для компьютерной лингводидактики.

7. Контент-анализ.

Все многообразие методик компьютерного анализа текста можно представить следующими направлениями: 1. Фоносемантический анализ – оценка звучания текста без учета его содержания, с использованием биполярных шкал (хороший / плохой, красивый / отталкивающий, безопасный / страшный и др.) и других механизмов. Подобный анализ позволяет определить возможный эффект воздействия текста на массовую аудиторию. Интересным примером использования фоносемантического анализа для выявления социальных рисков деятельности СМИ является анализ статей о насилии и массовых убийствах в российских учебных заведениях, позволивший выявить эмоциональное восприятие людьми фамилий убийц и заголовков статей, публикуемых в СМИ. Примеры организации фоносемантического анализа текста: – пример автоматизации фоносемантического анализа текста (программа ВААЛ(R) – <http://www.vaal.ru>), – пример автоматизации фоносемантического анализа слова (программа PsiTechnology, on-line версия – <https://psitechnology.net/servisfonosemantika.php>), – пример автоматизации фоносемантического анализа писем (программа АнализПисем.RU, on-line версия – <http://www.analizpisem.ru/index.html>). Анализ настроения автора письма осуществляется на основе следующих принципов: буквы русского языка обладают определенной частотой встречаемости, звуки человеческой речи имеют определенное подсознательное значение (Ч. Осгуд, А.П. Журавлёв), определяемое при помощи биполярных шкал (способных оценивать влияние звуков на эмоциональное состояние человека). Если



автор письма употребляет определенные буквы чаще или реже их средней частоты встречаемости, то эти звуки обладают для этого человека определённой подсознательной значимостью, что отражает настроения автора письма. 2. Контент-анализ – количественный анализ, ориентированный на оценку частотного распределения слов. Для простого подсчета частоты слов в тексте существует много программ (on-line версия – <https://webscript.ru/cgi-bin/text/text.cgi>, on-line версия – <https://gsgen.ru/tools/dlina-seo-text/> и др.) 3. Морфологический анализ – определяет морфологические интерпретации слов текста: – пример автоматизации морфологического анализа слов (on-line версия – <https://starling.rinet.ru/morph.htm>) 4. Синтаксический анализ направлен на получение синтаксической модели предложения, представленной в виде дерева зависимостей: – пример автоматизации синтаксического анализа предложения (программа АОТ (Автоматическая Обработка Текста), on-line версия – <http://www.aot.ru/onlinedemo.html>) 5. Семантический анализ – определяет семантическую структуру предложений. Программы, осуществляющие семантический анализ текстов: – пример автоматизации семантического анализа текста (on-line версия – https://miratext.ru/seo_analiz_text) Программ, осуществляющих семантический анализ, достаточно много: on-line версия программы Text.ru (<https://text.ru/>), on-line версия программы Istio (<https://istio.com/rus/text/analyz/>), on-line версия программы SEOLib (<https://seolib.ru/tools/text/analysis/>), on-line версия программы INtexty (<https://intexty.com/#seo-analysis-results>), программа TextAnalyst 2.0 (<https://www.analyst.ru/index.php?lang=eng&dir=content/products/&id=ta>) и др. 6. Интент-анализ – экспертное оценивание текста с позиции определения направленности автора на объект (интенции). 7. Дискурс-анализ – методики интерпретации текстов как продуктов речи, как пример социальной коммуникации. 8. Нарративный анализ – анализ текстов, содержащих элементы рассказов, ориентированный на оценку таких категорий как субъект, действия и др. 9. Экспертная оценка текста направлена на информацию об авторах текста, условиях его написания, о достоверности сведений и других характеристиках. 10. Графематический анализ используется для разделения текста на слова, поиска и выделения фамилий, дат, числовых данных и др. Кроме данных методик существует множество авторских разработок автоматизации процедур обработки текстов: Худломер (on-line версия – <http://teneta.rinet.ru/hudlomer/>) – стилистический анализ,

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 38	Первый экземпляр _____	КОПИЯ № _____

позволяет осуществлять классификацию функционального стиля текста на основе анализа длин слов. Анализ распределения длин слов в тексте позволяет характеризовать тексты с позиции их принадлежности к функциональным стилям (научный, информационно-новостной, нарратив / повествовательная проза, разговорная речь).

4. ПОРЯДОК ПРОВЕДЕНИЯ И КРИТЕРИИ ОЦЕНИВАНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

4.1. Порядок проведения промежуточной аттестации

Промежуточная аттестация включает в себя опрос в устной форме. Опрос предполагает развернутые ответы на вопросы по всем разделам рабочей программы дисциплины.

4.2. Критерии оценивания промежуточной аттестации по видам оценочных средств.

Максимальный балл за работу — 100 баллов (с коэффициентом 30%).

4.2.1. Критерии оценивания результатов опроса (проводится в устной форме)

В опросе необходимо дать развернутый ответ на вопрос. За правильное выполнение задания обучающийся получает **100 баллов**.

Высокий уровень подготовки – ответ полный (**86-100 баллов**). Обучающийся последовательно излагает теоретический материал, демонстрирует высокую степень проработанности пройденной темы, приводит подробные классификации, иллюстрирует теоретические положения актуальным языковым материалом, умело использует терминологию, метаязык, обобщает языковые факты и самостоятельно делает выводы. Обучающийся способен аргументировано ответить на дополнительные вопросы, изложить свою точку зрения;

хороший уровень подготовки – ответ полный (**69-85 баллов**). Обучающийся последовательно излагает теоретический материал, но допускает неточности в использовании понятийного аппарата. Приводимые классификации и теоретические положения не всегда иллюстрируются языковыми примерами. Обучающийся использует терминологию, но не всегда верно идентифицирует используемые научные категории и явления.

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 39	Первый экземпляр _____	КОПИЯ № _____

Обучающийся в основном способен аргументировано ответить на дополнительные вопросы, изложить свою точку зрения;

удовлетворительный уровень подготовки – ответ неполный (**51-68 баллов**). Знания теоретического материала поверхностны, не подкреплены иллюстративным языковым материалом. Обучающийся с трудом отвечает на дополнительные вопросы и не всегда излагает свою точку зрения;

неудовлетворительный уровень подготовки – ответ неполный (**10 баллов**) Выставляется, если ответ присутствует, но не удовлетворяет критериям знания материала, объема, логичности, наличия выводов и собственных примеров.

За отсутствие ответа выставляется **0 баллов**.

4.3. Результаты промежуточной аттестации и уровни сформированности компетенций

При подведении итогов учитываются результаты текущей аттестации. Полученные за текущую аттестацию баллы суммируются с баллами, полученными при прохождении промежуточной аттестации:

1. Текущая аттестация	70 %
1.1. Посещение занятий	5 %
1.2. Текущий контроль аудиторной работы	50%
1.3. Текущий контроль самостоятельной работы	15%
2. Промежуточная аттестация	30%
Итого:	100%
	= 100 баллов

Оценка	зачтено	зачтено	зачтено	незачтено
Баллы	100-90 баллов	89-80 баллов	79-69 баллов	68-0 баллов
Уровень освоения проверяемых компетенций	высокий	средний	базовый	низкий

Обучающийся, получивший в сумме от 51 до 100 баллов получает оценку «зачтено».

Обучающийся, получивший в сумме менее 51 баллов получает оценку «незачтено».

Особенности проведения процедуры оценивания результатов

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)			
Фонд оценочных средств для промежуточной аттестации по дисциплине «Современные технологии поиска и обработки информации» по направлению подготовки 45.04.02 Лингвистика направленности (профилю) Перевод и переводоведение ФГБОУ ВО «ЧелГУ»			
Версия документа - 1	стр. 40	Первый экземпляр _____	КОПИЯ № _____

обучения инвалидов и лиц с ограниченными возможностями здоровья обозначены в рабочей программе дисциплины (модуля).

Уровни сформированности компетенций по дисциплине определяется следующим образом:

1. Высокий уровень сформированности компетенций (100-90 баллов):

На высоком уровне владения компетенциями обучающийся демонстрирует всестороннее, систематическое и глубокое знание основных правил и принципов работы с электронными словарями и другими источниками информации в системе глобальных информационных ресурсов, а также правил и принципов представления информации с использованием информационных, компьютерных и сетевых технологий. Свободно и грамотно владеет основными способами поиска, хранения, обработки и анализа информации с использованием электронных словарей и других источников информации, оценки и анализа полученных данных посредством использования информационных ресурсов и технологий. Показывает глубокие знания основных правил и принципов работы с программными продуктами лингвистического профиля. Уверенно владеет стандартными методиками анализа, сравнения и рационального выбора современных информационных технологий для решения задач профессиональной деятельности. Умеет применять современные информационные технологии сбора, обработки и представления информации при проведении научных исследований в области лингвистики, межкультурной коммуникации и переводоведения; представлять информацию в требуемом формате с использованием информационных, компьютерных и сетевых технологий.

2. Средний уровень сформированности компетенций (89-80 баллов):

На среднем уровне обучающийся обладает уверенным знанием основных правил и принципов работы с электронными словарями и другими источниками информации в системе глобальных информационных ресурсов, а также правил и принципов представления информации с использованием информационных, компьютерных и сетевых технологий. Хорошо владеет основными способами поиска, хранения, обработки и анализа информации с использованием электронных словарей и других источников информации, оценки и анализа полученных данных посредством использования информационных ресурсов и технологий. Показывает достаточные знания основных правил и принципов работы с



программными продуктами лингвистического профиля. Хорошо владеет стандартными методиками анализа, сравнения и рационального выбора современных информационных технологий для решения задач профессиональной деятельности. Умеет применять современные информационные технологии сбора, обработки и представления информации при проведении научных исследований в области лингвистики, межкультурной коммуникации и переводоведения; представлять информацию в требуемом формате с использованием информационных, компьютерных и сетевых технологий.

Базовый уровень сформированности компетенций (79-69 баллов):

На базовом уровне обучающийся обладает достаточным знанием основных правил и принципов работы с электронными словарями и другими источниками информации в системе глобальных информационных ресурсов, а также правил и принципов представления информации с использованием информационных, компьютерных и сетевых технологий. Относительно хорошо владеет основными способами поиска, хранения, обработки и анализа информации с использованием электронных словарей и других источников информации, оценки и анализа полученных данных посредством использования информационных ресурсов и технологий. В отдельных случаях показывает знания основных правил и принципов работы с программными продуктами лингвистического профиля. Владеет некоторыми стандартными методиками анализа, сравнения и рационального выбора современных информационных технологий для решения задач профессиональной деятельности. В отдельных случаях умеет применять современные информационные технологии сбора, обработки и представления информации при проведении научных исследований в области лингвистики, межкультурной коммуникации и переводоведения; представлять информацию в требуемом формате с использованием информационных, компьютерных и сетевых технологий.

Низкий уровень сформированности компетенций (68-0 баллов).

Структура рабочей программы соответствует приказу ректора ФГБОУ ВО «ЧелГУ» от «13» апреля 2021 г. № 247-1