

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ИНСТИТУТ ПРИКЛАДНЫХ МАТЕМАТИЧЕСКИХ ИССЛЕДОВАНИЙ
КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ИПМИ КарНЦ РАН)**



УТВЕРЖДАЮ

Врио директора ИПМИ КарНЦ РАН

д.ф.-м.н., проф.

В.В. Мазалов

26 августа 2016 г.

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

ВЫЧИСЛИТЕЛЬНАЯ ЛИНГВИСТИКА

Направление подготовки: 09.06.01 Информатика и вычислительная техника

Профиль: Математическое моделирование, численные методы и комплексы программ

Квалификация: Исследователь. Преподаватель-исследователь

Петрозаводск
2016

Составитель рабочей программы:

Рук. лаб., к. т. н.

(должность, ученое звание, ученая степень)



(подпись)

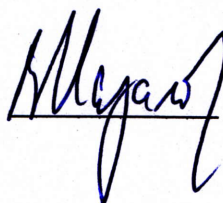
А.А. Крижановский

(Ф.И.О.)

Рабочая программа составлена в соответствии с ФГОС ВО (уровень подготовки кадров высшей квалификации) и утверждена на заседании Ученого совета ИПМИ КарНЦ РАН

«26» августа 2016 г., протокол № 9

Председатель Ученого совета
Д.ф.-м.н., проф.



В.В. Мазалов

1. Цели и задачи дисциплины

1.1. *Целью дисциплины является:* создание условий для формирования системы теоретических знаний в области вычислительной лингвистики, обработки текста на естественном языке; формирование навыков использования методов вычислительной лингвистики в научно-исследовательской и педагогической деятельности; формирование навыков разработки программных комплексов в научно-исследовательской деятельности; повышение квалификации в области научных основ и применении методов компьютерной лингвистики и комплексов программ для решения фундаментальных научных и прикладных научно-технических проблем.

1.2. *Виды профессиональной деятельности:*

- научно-исследовательская деятельность;
- проектная деятельность.

Выпускник, освоивший программу аспирантуры в соответствии с данными видами профессиональной деятельности, готов решать следующие профессиональные задачи:

- подготовка научных и научно-технических публикаций;
- разработка алгоритмов и программных комплексов с использованием методов вычислительной лингвистики;
- планирование процессов и ресурсов для решения задач в области обработки текста на естественном языке;
- использование методов вычислительной лингвистики в научно-исследовательской, педагогической и производственно-технологической деятельности, включая разработку алгоритмов и компьютерных программ для обработки и анализа текстов на естественном языке.

2. Место дисциплины в структуре ООП аспиранта

Дисциплина «Вычислительная лингвистика» является вариативной согласно учебному плану ООП по направлению подготовки 09.06.01 «Информатика и вычислительная техника», профиль – Математическое моделирование, численные методы и комплексы программ. Дисциплина изучается в 3-м и 4-м семестрах, направлена на формирование компетенций УК1, УК3, УК6, ОПК2, ОПК7, ПК1, ПК4.

3. Требования к уровню подготовки аспиранта, завершившего изучение дисциплины

В результате освоения дисциплины аспирант приобретает следующие компетенции.

Компетенция	Код по ФГОС ВО (уровень подготовки кадров)	Структура компетенции	Дескрипторы (уровни) - основные признаки освоения (показатели достижения результата)		Формы и методы обучения, способствующие формированию и развитию компетенции
Способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях	УК1	Знать: текущее состояние современных научных достижений, фундаментальные основы вычислительной лингвистики	Высокий (превосходный) уровень	Знать: основные понятия, модели, алгоритмы и теоретические положения курса «Вычислительная лингвистика»	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	основные задачи и алгоритмы вычислительной лингвистики	
			Пороговый (базовый) уровень	основные принципы математического моделирования и программирования	
		Уметь: генерировать новые идеи при решении исследовательских и практических задач, в том числе в междисциплинарных областях	Высокий (превосходный) уровень	Уметь: генерировать новые идеи при решении исследовательских и практических задач	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	эффективно использовать математические модели в научных исследованиях	
			Пороговый (базовый) уровень	решать конкретные практические задачи	
		Владеть: способностью к анализу и оценке современных научных	Высокий (превосходный) уровень	Владеть: основными методами научных исследований	Посещение лекций, семинаров, участие в научно-исследовательской

		достижений	Повышенный (продвинутый) уровень	навыками проведения вычислительного эксперимента	деятельности, применение полученных знаний для решения практических задач
			Пороговый (базовый) уровень	основными методами обработки данных с помощью программных комплексов	
Готовность участвовать в работе российских и международных исследовательских коллективов по решению научных и научно-образовательных задач	УКЗ	Знать: текущее состояние современных научных достижений	Высокий (превосходный) уровень	Знать: основные понятия, методы и алгоритмы курса «Вычислительная лингвистика»	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	основные задачи и методы их решения в вычислительной лингвистике	
			Пороговый (базовый) уровень	основные задачи вычислительной лингвистики, современные основы программирования	
		Уметь: вести научно-исследовательскую деятельность	Высокий (превосходный) уровень	Уметь: применять полученные теоретические знания для решения новых практических задач	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение знаний для решения практических задач
			Повышенный (продвинутый) уровень	эффективно использовать математические модели в научных исследованиях	
			Пороговый (базовый) уровень	решать практические задачи	
		Владеть: организационными, коммуникативными	Высокий (превосходный) уровень	Владеть: основными методами научных исследований	Посещение лекций, семинаров, участие в научно-исследовательской

		навыками, позволяющими осуществлять работу в исследовательских коллективах	Повышенный (продвинутый) уровень	навыками проведения вычислительного эксперимента	деятельности, применение полученных знаний для решения практических задач
			Пороговый (базовый) уровень	методикой решения практических задач	
Способность планировать и решать задачи собственного профессионального и личностного развития	УК6	Знать: текущее состояние современных научных достижений	Высокий (превосходный) уровень	Знать: основные понятия, методы и алгоритмы курса «Вычислительная лингвистика»	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	основные задачи и методы их решения в вычислительной лингвистике	
			Пороговый (базовый) уровень	основные задачи вычислительной лингвистики, современные основы программирования	
		Уметь: принимать мотивированное решение	Высокий (превосходный) уровень	Уметь: применять полученную теоретическую подготовку для постановки и решения практических задач	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение знаний для решения практических задач
			Повышенный (продвинутый) уровень	эффективно использовать математические модели в научных исследованиях	
			Пороговый (базовый) уровень	решать конкретные задачи	
		Владеть: навыками принятия решений и способностью нести	Высокий (превосходный) уровень	Владеть: основными методами научных исследований	Посещение лекций, семинаров, участие в научно-исследовательской

		ответственность за принятые решения	Повышенный (продвинутый) уровень	навыками математической обработки данных с помощью современных программных комплексов	деятельности, применение полученных знаний для решения практических задач
			Пороговый (базовый) уровень	навыками проведения вычислительного эксперимента	
Владение культурой научного исследования, в том числе с использованием современных информационно-коммуникационных технологий	ОПК2	Знать: текущее состояние современных научных достижений, методику проведения вычислительных экспериментов и разработки алгоритмов, реализации их на языках программирования	Высокий (превосходный) уровень	Знать: основные понятия, модели, алгоритмы и теоретические положения курса «Вычислительная лингвистика»	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	основные задачи и методы их решения в вычислительной лингвистике	
			Пороговый (базовый) уровень	основные задачи и методы их решения в вычислительной лингвистике, основные концепции программирования	
		Уметь: проводить вычислительные эксперименты, разрабатывать алгоритмы обработки текста, использовать компьютерные библиотеки и программные среды для обработки текста	Высокий (превосходный) уровень	Уметь: применять полученную теоретическую подготовку для решения новых практических задач	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	Успешно реализовывать алгоритмы в научных исследованиях	
			Пороговый (базовый) уровень	решать конкретные практические задачи	
		Владеть: навыками обработки информации и	Высокий (превосходный)	Владеть: основными методами научных исследова-	Посещение лекций, семинаров, участие в научно-

		анализа полученных данных, основными методами научных исследований, навыками проведения вычислительного эксперимента	уровень Повышенный (продвинутый) уровень Пороговый (базовый) уровень	ний навыками проведения вычислительного эксперимента навыками обработки экспериментальных данных с помощью программных комплексов	исследовательской деятельности, применение полученных знаний для решения практических задач
Владение методами проведения патентных исследований, лицензирования и защиты авторских прав при создании инновационных продуктов в области профессиональной деятельности	ОПК7	Знать: текущее состояние современных научных достижений, основы законодательства в области интеллектуальной собственности	Высокий (превосходный) уровень	Знать: теоретические положения курса «Вычислительная лингвистика», основы законодательства в области интеллектуальной собственности	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	основные задачи и методы их решения в вычислительной лингвистике; открытые лицензии, их сходство и разницу.	
			Пороговый (базовый) уровень	основные задачи и методы их решения в вычислительной лингвистике, основные концепции программирования	
		Уметь: применять полученные теоретические знания для решения конкретных практических задач, эффективно использовать математические модели в научных исследованиях	Высокий (превосходный) уровень Повышенный (продвинутый) уровень	Уметь: разрабатывать программные средства для обработки текста эффективно использовать существующие алгоритмы и разрабатывать новые для проведения научных исследований	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач

			Пороговый (базовый) уровень	решать конкретные практические задачи	
		Владеть: способностью к критическому анализу и оценке современных научных достижений	Высокий (превосходный) уровень	Владеть: основными методами научных исследований	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	навыками проведения вычислительного эксперимента	
			Пороговый (базовый) уровень	основными методами обработки данных с помощью современных программных комплексов	
Способность к разработке новых математических методов моделирования объектов и явлений	ПК1	Знать: современные подходы и алгоритмы, используемые при решении задач вычислительной лингвистики, методику проведения вычислительных экспериментов	Высокий (превосходный) уровень	Знать: основные теоретические положения курса «Вычислительная лингвистика»	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение знаний для решения практических задач
			Повышенный (продвинутый) уровень	основные алгоритмы, используемые при решении задач вычислительной лингвистики	
			Пороговый (базовый) уровень	основы классификации задач, методику решения задач и программирования	
		Уметь: проводить вычислительные эксперименты, разрабатывать математические модели и алгоритмы, использовать программные среды для решения задач вычислительной лингвистики	Высокий (превосходный) уровень	Уметь: применять полученные теоретические знания для решения новых практических задач	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	эффективно использовать алгоритмы в научных исследованиях	
			Пороговый (базовый) уровень	решать конкретные практические задачи	

			уровень		
		Владеть: навыками обработки информации и математического анализа полученных данных	Высокий (превосходный) уровень	Владеть: основными методами научных исследований	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	навыками проведения вычислительного эксперимента	
			Пороговый (базовый) уровень	основными методами обработки данных с помощью современных программных комплексов	
Готовность к реализации математического обеспечения в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента	ПК4	Знать: текущее положение современных научных достижений, методику проведения вычислительных экспериментов, современную методологию программирования	Высокий (превосходный) уровень	Знать: основные теоретические положения курса «Вычислительная лингвистика»	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение знаний для решения практических задач
			Повышенный (продвинутый) уровень	основные задачи и методы их решения в вычислительной лингвистике	
			Пороговый (базовый) уровень	основные задачи и методы их решения в вычислительной лингвистике, основные концепции программирования	
		Уметь: проводить вычислительные эксперименты, разрабатывать математические модели и алгоритмы, использовать программные среды для решения задач вычислительной лингвистики	Высокий (превосходный) уровень	Уметь: применять полученные теоретические знания для решения новых практических задач	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	эффективно использовать математические модели в научных исследованиях	
			Пороговый (базовый) уровень	решать конкретные практические задачи	

		Владеть: навыками обработки информации и математического анализа полученных данных	Высокий (превосходный) уровень	Владеть: основными методами научных исследований	Посещение лекций, семинаров, участие в научно-исследовательской деятельности, применение полученных знаний для решения практических задач
			Повышенный (продвинутый) уровень	навыками проведения вычислительного эксперимента	
			Пороговый (базовый) уровень	основными методами обработки данных с помощью современных программных комплексов	

4. Объем дисциплины и виды учебной работы

4.1. Объем дисциплины

Общая трудоемкость дисциплины составляет 2 зачетные единицы, 72 часа.

Вид учебной работы	Объем часов / зачетных единиц
Обязательная аудиторная учебная нагрузка (всего)	32
в том числе:	
лекции	14
семинары	-
практические занятия	18
Самостоятельная работа аспиранта (всего)	40
Вид контроля по дисциплине	зачет

4.2. Лекционные занятия

№ те мы	Название раздела/темы дисциплины	Технология проведения	Формиру- емые компетен- ции (код)	Форма оценочных средств	Трудоемкость (час.)
1	Вычислительная лингвистика	Чтение лекций, презентации с использованием мультимедийного оборудования, использование учебников, методических пособий и УМК	УК 3	коллоквиум	3
2	Информационный поиск	Чтение лекций, презентации с использованием мультимедийного оборудования, использование учебников, методических пособий и УМК	ПК 1	собеседование	3

3	Лексикография и корпуса	Чтение лекций, презентации с использованием мультимедийного оборудования, использование учебников, методических пособий и УМК	ОПК 1	коллоквиум	3
4	Корпусная лингвистика	Чтение лекций, презентации с использованием мультимедийного оборудования, использование учебников, методических пособий и УМК	ОПК 7	собеседование	3
5	Разрешение лексической многозначности	Чтение лекций, презентации с использованием мультимедийного оборудования, использование учебников, методических пособий и УМК	ПК 4	коллоквиум	2
Итого:					14

4.3. Практические (семинарские) занятия

№ темы	Тематика занятий	Технология проведения	Формируемые компетенции (код)	Форма оценочных средств	Трудоемкость (час.)
1	Основы работы в Викисловаре	Лабораторный практикум	УК 1	Дискуссия	4
2	Работа с внутренними ссылками, работа с шаблонами.	Лабораторный практикум	УК 6	Дискуссия	4
3	Работа в системе Викиданных	Лабораторный практикум, программирование	ОПК 1	Дискуссия	4
4	Открытый корпус русского языка.	Лабораторный практикум, задача наполнения корпуса	ОПК 7	Дискуссия	3
5	Лексикография и корпуса. Толковые	Лабораторный практикум,	ПК 1	Дискуссия	3

	словари.	создание словарных статей			
	Итого:				18

5. Содержание дисциплины

Тема №1. Вычислительная лингвистика (CL)

История развития лингвистики. Семиотика (Ч.С. Пирс, Ф. де Соссюр). Сравнительно-историческое языкознание. Разделы лингвистики. Смежные и пограничные науки (формальные грамматики, порождающие грамматики Н.Хомского). Квантитативная л-ка (з-н Мартина, Менцерата, Ципфа). CL и обработка текста на естественном языке. Три уровня лингвистического процессора. Задачи и направления CL.

Тема №2. Информационный поиск (ИП)

История ИП. Этапы развития сети Интернет. Определение, цель ИП. Классификация ИПС. Разница между Information Retrieval и Data Retrieval. Релевантность: информационная потребность и запрос, оценка ИПС, тестовая коллекция, стандартные тестовые коллекции. Pull & Push системы. Представление документа, нормализация текста, виды шумовых слов, нормализация слов. Архитектура ИПС.

Тема №3. Лексикография и корпуса

Лексикография и корпуса. Толковые словари.

Тема №4. Корпусная лингвистика

Корпусная лингвистика: определение, классификация, типы, примеры корпусов. Открытый корпус русского языка. Задача наполнения корпуса.

Тема №5. Разрешение лексической многозначности

Определение, классификация методов. Разрешение лексической многозначности в Викисловаре.

6. Учебно-методическое обеспечение самостоятельной работы аспирантов

Формы проведения самостоятельной работы

№	Тема дисциплины	Форма самостоятельной работы	Трудоемкость (час.)
1	Национальный корпус русского языка.	Задача наполнения корпуса. Разметка корпуса.	12
2	Лексикография и корпуса.	Изучение типов и требований к словарным статьям. Создание словарных статей в Викисловаре.	14
3	Экспертные системы и Викиданные	Знакомство с компьютерными программами, работающими с Викиданными.	14
	Итого:		40

7. Контроль знаний аспирантов

7.1. Формы текущего контроля работы аспирантов:

1) коллоквиум; 2) собеседование; 3) дискуссия.

7.2. Промежуточная аттестация по дисциплине:

проводится в форме зачета.

7.3. Вопросы по дисциплине

«Вычислительная лингвистика»

1. Вычислительная лингвистика. История развития лингвистики.
2. Семиотика (Ч.С. Пирс, Ф. де Соссюр). Сравнительно-историческое языкознание.
3. Разделы лингвистики. Смежные и пограничные науки (формальные грамматики, порождающие грамматики Н.Хомского). Квантитативная лингвистика (з-н Мартина, Менцерата, Ципфа).
4. Вычислительная лингвистика и обработка текста на естественном языке. Три уровня лингвистического процессора. Задачи и направления вычислительной лингвистики.
5. Вебометрия. Построение рейтинга вузов на основе стабильных свойств статей Википедии. Качество статей в Википедии (статус “добротных статей”).
6. Организация информации. Внутренние ссылки: когда нужны и когда излишни, два типа интервики, ссылки между проектами Викимедиа. Проект ВП «Связность»: определение тупиковых страниц и страниц-сирот, кластеры страниц.
7. Разрешение лексической многозначности. Определение, классификация методов. Разрешение лексической многозначности в Викисловаре.
8. Информационный поиск (ИП). История ИП. Этапы развития сети Интернет. Определение, цель ИП. Классификация информационно-поисковых систем (ИПС).
9. Релевантность: информационная потребность и запрос, оценка ИПС, тестовая коллекция, стандартные тестовые коллекции. Pull & Push системы. Представление документа, нормализация текста, виды шумовых слов, нормализация слов. Архитектура ИПС.
10. Инвертированный индекс. Модель булева поиска. Координатные и некоординатные инвертированные индексы. Этапы построения инвертированного индекса. Структура данных для списков словопозиций: односвязные списки, списки с пропусками, массивы переменной длины.
11. WikiData: проблемы Википедии, сравнение с Semantic MediaWiki, три задачи Викиданных.
12. Алгебра логики: булевы функции. Законы де Моргана, дизъюнктивная нормальная форма.
13. Расширенные модели булева поиска: оператор близости терминов, проблемы булевых запросов. Метрики точности, полноты, fall-out.
14. Оптимизация обработки булевых запросов. Пересечение списков словопозиций. Алгоритм слияния списков словопозиций.
15. Лексикография и корпуса. Толковые словари.
16. Корпусная лингвистика: определение, классификация, типы, примеры корпусов.
17. Открытый корпус русского языка. Задача наполнения корпуса.

7.4. Критерии оценки промежуточной аттестации аспирантов по дисциплине «Вычислительная лингвистика»

Критерии зачета с оценкой	
оценка «отлично»	Ставится, если аспирант строит ответ логично в соответствии с планом, обнаруживает глубокое знание теоретических вопросов. Уверенно отвечает на дополнительные вопросы. Грамотно использует научную лексику, свободно ориентируется в материале курса. Аспирант успешно справляется с практическим заданием.
оценка «хорошо»	Ставится, если аспирант строит ответ в соответствии с планом, обнаруживает понимание теоретических вопросов. Ответ содержит ряд несущественных неточностей. Наблюдается неточность при ответе на дополнительные вопросы. Аспирант успешно справляется с практическим заданием или допускает незначительные ошибки.
оценка «удовлетворительно»	Ставится, если ответ аспиранта недостаточно логически выстроен, обнаруживается недостаточно полное понимание теоретических вопросов, хотя основные понятия раскрываются правильно. Аспирант справляется с практическим заданием, допуская ошибки.
оценка «неудовлетворительно»	Ставится если, аспирант оказывается неспособным правильно раскрыть содержание основных понятий. Проявляет стремление подменить научное обоснование проблемы общими рассуждениями. Ответ содержит ряд серьезных неточностей. Аспирант не способен выполнить практическое задание.

Критерии оценки зачета	
«зачтено»	Ставится, если аспирант строит ответ логично в соответствии с планом, обнаруживает глубокое знание теоретических вопросов. Уверенно отвечает на дополнительные вопросы. При ответе грамотно использует научную лексику, свободно ориентируется в материале курса. Аспирант успешно справляется с практическим заданием.
«незачтено»	Ставится если, аспирант оказывается неспособным правильно раскрыть содержание основных понятий, плохо ориентируется в материале курса. Ответ содержит ряд серьезных неточностей. Аспирант не способен выполнить практическое задание.

7.5. Фонд оценочных средств

Содержание фонда оценочных средств: см. Приложение №1.

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Основная литература

1. Daniel Jurafsky & James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2nd Edition. Prentice-Hall. ISBN 978-0131873216. URL:

http://stp.lingfil.uu.se/~santini/ml/2014/JurafskyMartinSpeechAndLanguageProcessing2ed_draft%202007.pdf (draft).

2. Bird, Steven; Klein, Ewan; Loper, Edward (2009). Natural Language Processing with Python. O'Reilly Media Inc. ISBN 0-596-51649-5. URL: <http://www.nltk.org/book/>
3. Работа в вики-среде на примере Русской Википедии, часть 1 : учеб. пособие / А. А. Крижановский — Петрозаводск : Изд-во ПетрГУ, 2015. — 71 с. URL: <http://scipeople.com/publication/114999/>
4. Работа в вики-среде на примере Русской Википедии, часть 2 : учеб. пособие / А. А. Крижановский — Петрозаводск : Изд-во ПетрГУ, 2015. — 52 с. URL: <http://scipeople.com/publication/114999/>

8.2. Дополнительная литература

1. Прикладная и компьютерная лингвистика / (ред.) И. С. Николаев, О. В. Митренина, Т. М. Ландо. М., 2016. — 320 с. — ISBN 978-5-9710-3472-8.
2. Митрофанова О.А. Формальная семантика. СПб.: Факультет филологии и искусств СПбГУ, 2007.
3. Митрофанова О.А. Формальные методы в лингвистике. СПб.: Факультет филологии и искусств СПбГУ, 2007. 24 с.

8.3.

Интернет-ресурсы	
http://scholar.google.com	Академия Google, научные статьи
http://arxiv.org	Архив электронных публикаций научных статей
http://www.elibrary.ru	Российская научная электронная библиотека
http://ruscorpora.ru	Национальный корпус русского языка
http://opencorpora.org	Открытый корпус
https://ru.wiktionary.org	Русский Викисловарь
https://www.wikidata.org	Викиданные, совместно редактируемая база знаний.

9. Перечень программного обеспечения

Открытое (бесплатное) программное обеспечение: LaTeX, LibreOffice, MediaWiki, NLTK.

10. Материально-техническое обеспечение дисциплины

Аудитории для проведения лекционных и практических занятий, мультимедийное оборудование, доска, доступ к Интернет-ресурсам.

11. Особенности организации образовательного процесса для инвалидов и лиц с ограниченными возможностями здоровья

Обучение инвалидов и лиц с ограниченными возможностями осуществляется в соответствии со следующими документами.

1. Ст.79, 273-ФЗ «Об образовании в Российской Федерации» .
2. Раздел IV, п.п. 46-51 приказа Минобрнауки России от 19.11.2013 № 1259 «Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным

программам высшего образования – программам подготовки научно-педагогических кадров в аспирантуре (адъюнктуре)».

3. Методические рекомендации по организации образовательного процесса для обучения инвалидов и лиц с ограниченными возможностями здоровья в образовательных организациях высшего образования, в том числе оснащённости образовательного процесса (утверждены зам. Министра образования и науки РФ А.А.Климовым от 8.4.2014 г. № АК-44/05 вн).

Содержание фонда оценочных средств

Коллоквиум

Вопросы к коллоквиуму по дисциплине «Вычислительная лингвистика»:

1. Вычислительная лингвистика. Семиотика (Ч.С. Пирс, Ф. де Соссюр). Сравнительно-историческое языкознание.
3. Формальные грамматики, порождающие грамматики Н.Хомского. Квантитативная лингвистика (з-н Мартина, Менцерата, Ципфа).
4. Вычислительная лингвистика и обработка текста на естественном языке. Три уровня лингвистического процессора. Задачи и направления вычислительной лингвистики.
5. Вебометрия.
6. Организация информации. Внутренние ссылки.
7. Разрешение лексической многозначности. Определение, классификация методов.
8. Информационный поиск. Классификация информационно-поисковых систем.
9. Релевантность: информационная потребность и запрос, тестовая коллекция, стандартные тестовые коллекции. Представление документа, нормализация текста, нормализация слов.
10. Инвертированный индекс. Модель булева поиска. Координатные и некоординатные инвертированные индексы. Этапы построения инвертированного индекса. Структура данных для списков словопозиций: односвязные списки, списки с пропусками, массивы переменной длины.
11. WikiData: проблемы Википедии, сравнение с Semantic MediaWiki, три задачи Викиданных.
12. Алгебра логики: булевы функции. Законы де Моргана, дизъюнктивная нормальная форма.
13. Расширенные модели булева поиска: оператор близости терминов, проблемы булевых запросов. Метрики точности, полноты, fall-out.
14. Оптимизация обработки булевых запросов. Пересечение списков словопозиций. Алгоритм слияния списков словопозиций.
15. Лексикография и корпуса. Толковые словари.
16. Корпусная лингвистика: определение, классификация, типы, примеры корпусов.
17. Открытый корпус русского языка. Разметка корпуса.

Критерии оценки коллоквиума

«зачтено»	Ставится, если аспирант строит ответ логично в соответствии с планом, обнаруживает глубокое знание теоретических вопросов. Уверенно отвечает на дополнительные вопросы.
«незачтено»	Ставится, если аспирант оказывается неспособным правильно раскрыть содержание основных понятий. Ответ содержит ряд серьезных неточностей. Аспирант не отвечает на дополнительные вопросы.

Собеседование

Тема №1. Национальный корпус русского языка

Поисковая система корпуса. Выбор и построение подкорпуса. Поиск в основном корпусе. Распределение употреблений слов по годам.

Тема №2. Лексикография и корпуса

Структура словарной статьи. Правила оформления словарных статей в Викисловаре.

Тема №3. Экспертные системы и Викиданные

Викиданные. Назначение системы, решаемые задачи. Структура данных.

Компьютерная программа — бот. Возможности ботов. Примеры ботов. Области их использования.

Критерии оценки собеседования

«зачтено»	Ставится, если аспирант строит ответ логично в соответствии с планом, обнаруживает глубокое знание теоретических вопросов. Уверенно отвечает на дополнительные вопросы. При ответе грамотно использует научную лексику, способен привести примеры, демонстрирующие эффективность теории.
«незачтено»	Ставится, если аспирант оказывается неспособным правильно раскрыть содержание основных понятий. Ответ содержит ряд серьезных неточностей. Аспирант не отвечает на дополнительные вопросы и не ориентируется свободно в излагаемом материале.

Дискуссия

Темы дискуссий

1. Вычислительная лингвистика. Семиотика. Сравнительно-историческое языкознание.
2. Формальные грамматики, порождающие грамматики Н.Хомского.
3. Квантитативная лингвистика (з-н Мартина, Менцерата, Ципфа).
4. Вебометрия.
5. Разрешение лексической многозначности. Определение, классификация методов.
6. Информационный поиск. Классификация информационно-поисковых систем.

7. Релевантность: информационная потребность и запрос, тестовая коллекция, стандартные тестовые коллекции. Представление документа, нормализация текста, нормализация слов.
8. Инвертированный индекс. Модель булева поиска.
9. Викиданные и экспертные системы.
10. Расширенные модели булева поиска: оператор близости терминов, проблемы булевых запросов. Метрики точности, полноты, fall-out.
11. Лексикография и корпуса.
12. Корпусная лингвистика.
- 17. Открытый корпус русского языка. Разметка корпуса.**

Критерии оценки дискуссии:

«зачтено»	Ставится, если аспирант раскрывает тему дискуссии логично, обнаруживает глубокое знание темы. Уверенно отвечает на вопросы, грамотно обосновывает свою позицию. При ответе свободно и уверенно ориентируется в материале.
«незачтено»	Ставится, если аспирант оказывается неспособным правильно раскрыть содержание основных понятий по теме дискуссии. Ответ содержит ряд серьезных неточностей. Аспирант не раскрывает тему дискуссии, не обосновывает свою позицию по теме дискуссии.

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ В РАБОЧЕЙ ПРОГРАММЕ

за _____ / _____ учебный год

В рабочую программу _____
(наименование дисциплины)

Для специальности (тей) _____
(номер специальности)

Вносятся следующие дополнения и изменения:

Дополнения и изменения внес _____
(должность, ФИО, подпись)

Рабочая программа пересмотрена и одобрена на заседании Ученого совета ИПМИ КарНЦ
РАН

« ____ » _____ 20 ____ г.

Председатель Ученого совета _____
(подпись) (ФИО)