

Минобрнауки России  
Федеральное государственное  
бюджетное учреждение науки  
**Федеральный исследовательский центр**  
**«Карельский научный центр**  
**Российской академии наук»**  
(КарНЦ РАН)

**УТВЕРЖДАЮ**

Врио председателя КарНЦ РАН  
член-корр. РАН

\_\_\_\_\_ О.Н. Бахмет

« \_\_\_\_ » \_\_\_\_\_ 2018 г.

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

### **«Вычислительная лингвистика»**

Основной образовательной программы высшего образования –  
программы подготовки научно-педагогических кадров в аспирантуре  
по направлению подготовки

**09.06.01 Информатика и вычислительная техника,**  
**профиль: Математическое моделирование,**  
**численные методы и комплексы программ**

Петрозаводск  
2018

Рабочая программа дисциплины разработана в соответствии с ФГОС ВО, утвержденным приказом Министерства образования и науки Российской Федерации от 30.07.2014 г. № 875 и учебным планом по направлению подготовки аспирантуры 09.06.01 Информатика и вычислительная техника, профиль «Математическое моделирование, численные методы и комплексы программ».

Разработчик:

Крижановский Андрей Анатольевич, ведущий научный сотрудник ИПМИ КарНЦ РАН, кандидат технических наук

---

Эксперт:

Реттеева Анна Николаевна, ведущий научный сотрудник ИПМИ КарНЦ РАН, доктор физико-математических наук, доцент

---

Рабочая программа дисциплины рассмотрена и одобрена на заседании Ученого совета ИПМИ КарНЦ РАН (протокол № 5 от «24» мая 2018 г.) и рекомендована к утверждению на заседании Ученого совета КарНЦ РАН (протокол № 7 от «25» мая 2018 г.)

**1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения основной образовательной программы (ООП).**

Компетенции обучающегося, формируемые в результате освоения дисциплины:

Код компетенции. Этап формирования компетенции	Формулировка компетенции	Планируемые результаты обучения (индикаторы достижения компетенции)
УК1	Способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях	<p><b>Знать:</b> текущее состояние современных научных достижений, фундаментальные основы вычислительной лингвистики</p> <p><b>Уметь:</b> генерировать новые идеи при решении исследовательских и практических задач, в том числе в междисциплинарных областях.</p> <p><b>Владеть навыками (опытом деятельности):</b> способностью к анализу и оценке современных научных достижений.</p>
УК3	Готовность участвовать в работе российских и международных исследовательских коллективов по решению научных и научно-образовательных задач	<p><b>Знать:</b> текущее состояние современных научных достижений.</p> <p><b>Уметь:</b> вести научно-исследовательскую деятельность.</p> <p><b>Владеть навыками (опытом деятельности):</b> организационными, коммуникативными навыками, позволяющими осуществлять работу в исследовательских коллективах.</p>
УК6	Способность планировать и решать задачи собственного профессионального и личностного развития	<p><b>Знать:</b> текущее состояние современных научных достижений.</p> <p><b>Уметь:</b> принимать мотивированное решение.</p> <p><b>Владеть навыками (опытом деятельности):</b> навыками принятия решений и способностью нести ответственность за принятые решения.</p>
ОПК2	Владение культурой научного исследования, в том числе с использованием современных	<p><b>Знать:</b> текущее состояние современных научных достижений, методику проведения вычислительных экспериментов и разработки алгоритмов, реализации их на языках программирования.</p>

	информационно-коммуникационных технологий	<p><b>Уметь:</b> проводить вычислительные эксперименты, разрабатывать алгоритмы обработки текста, использовать компьютерные библиотеки и программные среды для обработки текста.</p> <p><b>Владеть навыками (опытом деятельности):</b> навыками обработки информации и анализа полученных данных, основными методами научных исследований, навыками проведения вычислительного эксперимента.</p>
ОПК7	Владение методами проведения патентных исследований, лицензирования и защиты авторских прав при создании инновационных продуктов в области профессиональной деятельности	<p><b>Знать:</b> текущее состояние современных научных достижений, основы законодательства в области интеллектуальной собственности.</p> <p><b>Уметь:</b> применять полученные теоретические знания для решения конкретных практических задач, эффективно использовать математические модели в научных исследованиях.</p> <p><b>Владеть навыками (опытом деятельности):</b> способностью к критическому анализу и оценке современных научных достижений.</p>
ПК1	Проектирование и разработка подходов, алгоритмов и приложений, умение проведения вычислительных экспериментов.	<p><b>Знать:</b> современные подходы и алгоритмы, используемые при решении задач вычислительной лингвистики, методику проведения вычислительных экспериментов .</p> <p><b>Уметь:</b> проводить вычислительные эксперименты, разрабатывать математические модели и алгоритмы, использовать программные среды для решения задач вычислительной лингвистики.</p> <p><b>Владеть навыками (опытом деятельности):</b> навыками обработки информации и математического анализа полученных данных.</p>
ПК4	Готовность к реализации математического обеспечения в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента	<p><b>Знать:</b> текущее положение современных научных достижений, методику проведения вычислительных экспериментов, современную методологию программирования .</p> <p><b>Уметь:</b> проводить вычислительные эксперименты, разрабатывать математические модели и алгоритмы, использовать программные среды для решения задач вычислительной лингвистики.</p>

		<b>Владеть навыками (опытом деятельности):</b> навыками обработки информации и математического анализа полученных данных.
--	--	--

## 2. Место дисциплины в структуре ООП аспирантуры и язык преподавания

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 и относится к дисциплинам по выбору.

Дисциплина изучается в 3-м и 4-м семестрах, направлена на формирование компетенций УК1, УК3, УК6, ОПК2, ОПК7, ПК1, ПК4.

Язык преподавания – русский.

## 3. Виды учебной работы и тематическое содержание дисциплины

Общая трудоемкость дисциплины составляет 2 зачетные единицы или 72 академических часов.

### 3.1. Виды учебной работы

Виды учебной работы	Объем в академических часах
<b>Общая трудоемкость дисциплины по учебному плану</b>	72
В том числе:	
<b>Контактная работа (работа во взаимодействии с преподавателем). Всего</b>	32
В том числе:	
Лекции (Л)	14
Практические занятия (Пр)	18
Лабораторные занятия (Лаб)	0
Вид промежуточной аттестации	зачет
<b>Самостоятельная работа обучающихся (СР) (всего)</b>	40
В том числе:	
Самостоятельное изучение разделов дисциплины, подготовка к занятиям	36
Подготовка к промежуточной аттестации	4

### 3.2. Краткое содержание дисциплины по разделам и видам учебной работы

№ п/п	Раздел дисциплины (тематический модуль)	Трудоемкость по видам учебных занятий (в академических часах)					Оценочное средство
		Всего	Лекции	Практические занятия	Лабораторные занятия	Самостоятельная работа обучающихся	
<b>Семестр № 3,4</b>							
1	Вычислительная лингвистика. История развития лингвистики.		3	4	0	12	коллоквиум
2	Разделы лингвистики. Смежные и пограничные науки (формальные грамматики, порождающие грамматики Н.Хомского).		3	3	0	0	собеседование
3	Квантитативная лингвистика (3-н Мартина, Менцерата, Ципфа).		3	4	0	14	коллоквиум
4	Вычислительная лингвистика и обработка текста на естественном языке. Три уровня лингвистического процессора. Задачи и направления вычислительной лингвистики.		3	4		0	собеседование
5	Разрешение лексической многозначности. Лексикография и корпуса. Толковые словари.		2	3		14	коллоквиум
Вид промежуточной аттестации в семестре: зачет (зачёт и консультация: 4 часа)							
<b>Итого:</b>		32	14	18	0	40	

### 3.3. Содержание аудиторных занятий

#### Содержание лекционных занятий

№ раздела	№ лекции	Основное содержание	Количество часов	В т.ч. с использованием ДОТ (*)
<b>Семестр № 3</b>				
1	1.1	Вычислительная лингвистика (CL). История развития лингвистики. CL и обработка текста на естественном языке. Задачи и направления CL.	1	0

2	2.1	Семиотика. Сравнительно-историческое языкознание.	1	0
3	3.1	Разделы лингвистики. Смежные и пограничные науки (формальные грамматики, порождающие грамматики Н.Хомского).	1	0
4	4.1	Квантитативная лингвистика (з-н Мартина, Менцерата, Ципфа).	1	0
5	5.1	Вычислительная лингвистика и обработка текста на естественном языке. Три уровня лингвистического процессора. Задачи и направления вычислительной лингвистики.	1	0
6	6.1	Викиданные и вики-технологии: принципы работы, проблемы Википедии, сравнение с Semantic MediaWiki. Использование данных Википедии в вычислительной лингвистике.	1	0
7	7.1	Нормализация текста.	1	0
8	8.1	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.	1	0
9	9.1	Корпусная лингвистика: определение, классификация, типы, примеры корпусов. Лексикография и корпуса. Толковые словари. Электронная лексикография.	1	0
10	10.1	Языки визуального программирования: Scratch и App Inventor. Настройка среды App Inventor. Галерея App Inventor: Публикация приложений.	1	1
10	10.2	App Inventor: Переменные (глобальные и локальные). Процедуры и аргументы.	1	1
10	10.3	App Inventor: Менеджер экранов, работа со списками.	1	1
10	10.4	Проектирование и разработка приложения в среде App Inventor.	1	1
11	11.1	Система компьютерной вёрстки LaTeX.	1	
<b>Итого:</b>			<b>14</b>	<b>4</b>

### Содержание практических занятий

№ раздела	№ занятия	Основное содержание	Количество часов	В т.ч. с использованием ДОТ (*)
Семестр № 4				
11	1	Выбор задачи.	1	0
11-12	2	Установка системы вёрстки LaTeX и настройка системы совместной разработки GitHub.	1	0
11	3	Создание первого окна приложения About (об авторе, благодарности и о приложении).	1	0
11	4	Разработка математической задачи для последующей реализации в приложении.	1	0
11	5-7	Программирование в среде App Inventor	3	2
11	8-10	Иллюстрирование статьи.	3	2
12	11	Оформление ссылок на литературу и веб-ресурсы.	1	0

12	12	Индексирование ключевых слов в системе LaTeX.	1	0
12	13	Подготовка Аннотации	1	0
11	14-15	Разработка игры в форме упражнений (quiz). Вопросы по программированию в среде App Inventor и по содержанию задачи (приложения).	3	2
11	16	Записать видео с примером работы программы.	1	1
12	17	Компиляция LaTeX в PDF	1	0
<b>Итого</b>			<b>18</b>	<b>7</b>

#### 4. Образовательные технологии по дисциплине

Реализация компетентного подхода предусматривает широкое использование в учебном процессе активных и интерактивных форм проведения занятий в сочетании с внеаудиторной работой. В рамках освоения дисциплины проводятся обсуждения, обучающиеся готовят презентации и выступают с докладами по темам, связанным с вычислительной лингвистикой и обработкой текста. Используются такие образовательные технологии: разбор конкретных ситуаций и тренинги.

Студенты изучают и представляют результаты работы с помощью вики-проектов. Подробнее это описано в публикации в СМИ [1], [2] и в учебном пособии [3], [4].

[1]. Krizhanovsky A. Russia report: WikiProjects of students of Petrozavodsk State University // This Month in GLAM – Volume VII, Issue VI, June 2017. URL: [https://outreach.wikimedia.org/wiki/GLAM/Newsletter/June\\_2017/Contents/Russia\\_report](https://outreach.wikimedia.org/wiki/GLAM/Newsletter/June_2017/Contents/Russia_report)

[2]. A. Krizhanovsky, The first students of Russia with diplomas of Wikimedia and Petrozavodsk State University // This Month in Education – Volume 4, Issue 5, May 2018. URL: [https://outreach.wikimedia.org/wiki/Education/News/May\\_2018/The\\_first\\_students\\_of\\_Russia\\_with\\_diplomas\\_of\\_Wikimedia\\_and\\_Petrozavodsk\\_State\\_University](https://outreach.wikimedia.org/wiki/Education/News/May_2018/The_first_students_of_Russia_with_diplomas_of_Wikimedia_and_Petrozavodsk_State_University)

[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://nauchkor.ru/uploads/documents/5690f7f35f1be74d9400018e.pdf>

[4]. Работа в вики-среде на примере Русской Википедии, часть 2 : учеб. пособие / А. А. Крижановский — Петрозаводск : Изд-во ПетрГУ, 2015. — 52 с.

#### 5. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

5.1. Текущий контроль осуществляется преподавателем дисциплины при проведении занятий в форме:

- докладов
- лабораторных работ
- проекта (написание научно-исследовательской статьи в Викиверситете).

Список из 12 контрольных вопросов и заданий для проведения текущего контроля доступен на странице [https://ru.wikiversity.org/wiki/App\\_Inventor](https://ru.wikiversity.org/wiki/App_Inventor).

5.2. Промежуточная аттестация проводится в виде зачета.

## Вопросы к зачету

1. Лицензии, авторское право и Википедия.
2. Основы работы в Википедии.
3. Информационный поиск.
4. Вебметрика.
5. Работа в системе Викиданных.
6. Нормализация текста.
7. 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.
8. Корпусная лингвистика: определение, классификация, типы, примеры корпусов.
9. Открытый корпус русского языка. Задача наполнения корпуса.
10. Языки визуального программирования: Scratch и App Inventor.
11. Особенности публикации приложений в Галерее App Inventor.
12. Переменные в среде App Inventor.
13. Процедуры и аргументы в среде App Inventor.
14. Менеджер экранов в среде App Inventor.
15. Работа со списками в среде App Inventor.
16. Система компьютерной вёрстки LaTeX.

Подробно средства оценивания для проведения промежуточной аттестации обучающихся приведены в Фонде оценочных средств по данной дисциплине.

## 6. Методические рекомендации обучающимся по дисциплине, в том числе для самостоятельной работы

### Формы проведения самостоятельной работы

№	Тема дисциплины	Форма самостоятельной работы	Трудоемкость (час.)
1	Основы обработки текстов.	Нормализация текста. Информационный поиск.	12
2	Лексикография и корпуса.	Решение задач на основе Открытого корпуса русского языка.	14
3	Экспертные системы и Викиданные	Исследование возможностей Викиданных.	14
	Итого:		40

Студентам рекомендуется ознакомиться со следующими материалами и учебным пособием:

[1]. Krizhanovsky A. Russia report: WikiProjects of students of Petrozavodsk State University // This Month in GLAM – Volume VII, Issue VI, June 2017. URL: [https://outreach.wikimedia.org/wiki/GLAM/Newsletter/June\\_2017/Contents/Russia\\_report](https://outreach.wikimedia.org/wiki/GLAM/Newsletter/June_2017/Contents/Russia_report)

[2]. A. Krizhanovsky, The first students of Russia with diplomas of Wikimedia and Petrozavodsk State University // This Month in Education – Volume 4, Issue 5, May 2018. URL: [https://outreach.wikimedia.org/wiki/Education/News/May\\_2018/The\\_first\\_students\\_of\\_Russia\\_with\\_dipl](https://outreach.wikimedia.org/wiki/Education/News/May_2018/The_first_students_of_Russia_with_dipl)

[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.

[5]. Онлайн-курс "App Inventor", разработанный Крижановским А.А. [https://ru.wikiversity.org/wiki/App\\_Inventor](https://ru.wikiversity.org/wiki/App_Inventor)

Рекомендации обучающимся при освоении лекционного материала:

- конспектирование основного содержания лекций;
- для лучшего усвоения материала после лекции рекомендуется прочесть конспект и соответствующий параграф или главу учебника.

## **7. Методические рекомендации преподавателям по дисциплине**

Преподавателям рекомендуется использовать материалы, доступные на странице <http://mathem.krc.karelia.ru/section.php?plang=r&id=925>

## **8. Учебно-методическое и информационное обеспечение дисциплины**

Дисциплина полностью обеспечена учебной литературой, представленной в печатном или электронном виде. Для осуществления образовательной деятельности по дисциплине рекомендуется следующая основная и дополнительная литература.

### **8.1. Основная литература:**

1. 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/>
2. 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: <https://web.stanford.edu/~jurafsky/slp3/>
3. 3. Работа в вики-среде на примере Русской Википедии, часть 1 : учеб. пособие / А. А. Крижановский — Петрозаводск : Изд-во ПетрГУ, 2015. — 71 с. URL: <https://nauchkor.ru/uploads/documents/5690f7f35f1be74d9400018e.pdf>

### **8.2. Дополнительная литература:**

1. Baeza-Yates R., Ribeiro-Neto B. Modern Information Retrieval. — Addison-Wesley, 1999. URL: <http://people.ischool.berkeley.edu/~hearst/irbook/>
2. Manning, C.D. Introduction to Information Retrieval [Электронный ресурс] / C.D. Manning, P. Raghavan and H. Schütze. - Cambridge University Press, 2008. - URL: <https://nlp.stanford.edu/IR-book/information-retrieval-book.html>. Яз. англ. (10.05.2018)
3. Математическая лингвистика и автоматическая обработка текстов [Электронный ресурс]: учеб. пособие / Батура Т.В. - М. : Новосиб. гос. ун-т. - Ново-сибирск : РИЦ НГУ, 2016. - 166 с. - ISBN 978-5-4437-0548-4. - URL: <http://www.studentlibrary.ru/book/ISBN9785443705484.html>. Яз. рус. (10.05.2018)

### 8.3. Программное обеспечение и Интернет-ресурсы:

1. Электронно-библиотечная поисковая система «Академия Google». <http://scholar.google.com/>
2. Викиданные, совместно редактируемая база знаний <https://www.wikidata.org>
3. Национальный корпус русского языка [ruscorpora.ru](http://ruscorpora.ru)

### 8.4. Информационное обеспечение дисциплины в системе электронного (дистанционного) обучения

- Онлайн-курс "App Inventor", разработанный Крижановским А.А. на платформе Викиверситета. 2018. URL: [https://ru.wikiversity.org/wiki/App\\_Inventor](https://ru.wikiversity.org/wiki/App_Inventor)
- Крижановский А.А. Презентации по спецкурсу "Интернет-математика / Спецглавы математики". 2012-2018. URL: <http://mathem.krc.karelia.ru/section.php?plang=r&id=925>

## 9. Материально-техническое обеспечение дисциплины

Материально-техническая база КарНЦ обеспечивает проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работы обучающихся, предусмотренных учебным планом и соответствует действующим санитарным и противопожарным правилам и нормам.

Минимально-необходимый перечень для информационно-технического и материально-технического обеспечения дисциплины:

- аудитория для проведения лекционных и практических занятий, оснащенная рабочими местами для обучающихся и преподавателя, доской, мультимедийным оборудованием;
- библиотека с читальным залом и залом для самостоятельной работы обучающегося, оснащенное компьютером с выходом в Интернет, книжный фонд которой составляет специализированная научная, учебная и методическая литература, журналы (в печатном или электронном виде).