

24  
ТВУиНД

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Петрозаводский государственный университет

Кафедра теории вероятностей и анализа данных

УТВЕРЖДАЮ

декан математического факультета

А. Г. Варфоломеев

« 25 » 06 2012 г.



Рабочая программа дисциплины

**Интернет-математика**

010100 Математика

Квалификация (степень) выпускника:

Магистр

Петрозаводск

2012 г.

### **Общие сведения о дисциплине**

Название дисциплины – **«Интернет-математика»**

Факультет, на котором преподается данная дисциплина – **Математический**

Направление подготовки – **010100 Математика**

Квалификация (степень) выпускника – **Магистр**

Цикл дисциплин – **Профессиональный**

Часть цикла – **Вариативная**

Курс – **2**

Семестры – **3**

Всего зачетных единиц – **5**

Всего часов – **180**

Аудиторные занятия **34 часа** (лекции – **17 часов**, лабораторные занятия – нет, практические занятия – **17 часов**)

Объем часов в активной и интерактивной формах – **10 часов**

Самостоятельная работа – **146 часов**

Экзамен – **3 семестр**

Составитель рабочей программы – старший преподаватель, кандидат технических наук **Крижановский Андрей Анатольевич**

## **1. Цели освоения дисциплины**

Целями освоения дисциплины «Интернет-математика» являются: создание условий для формирования системы теоретических знаний в области информационного поиска, компьютерной лингвистики, обработки текста на естественном языке, теории графов и семантического веба.

Задачами курса являются изучение основных проблем, решаемых вычислительной лингвистикой и овладение навыками работы с лингвистическими программами, поисковыми системами и сервисами сети Интернет.

## **2. Место дисциплины в структуре ООП магистратуры**

Дисциплина «Интернет-математика» относится к профессиональному циклу дисциплин, вариативной части ФГОС по направлению подготовки «Математика». Для изучения дисциплины необходимы компетенции, сформированные у студентов в результате обучения в образовательных учреждениях высшего профессионального образования по программам бакалавриата.

Овладение основами курса способствует подготовке студентов к успешной работе в различных сферах, применяющих математические методы на основе гармоничного сочетания научной, фундаментальной и профессиональной подготовки кадров, формированию социально-личностных качеств и развитию способности самостоятельно приобретать и применять новые знания и умения.

## **3. Компетенции обучающегося, формируемые в результате освоения дисциплины «Интернет-математика».**

### **Общекультурные:**

- способность работать самостоятельно, заботиться о качестве, стремиться к успеху (ОК-6);
- умение находить, анализировать и контекстно обрабатывать информацию, в том числе относящуюся к новым областям знаний, непосредственно не связанным со сферой профессиональной деятельности (ОК-10).

### **Профессиональные компетенции:**

- владение методами математического моделирования при анализе глобальных проблем на основе глубоких знаний фундаментальных математических дисциплин и компьютерных наук (ПК-1);
- владение методами математического и алгоритмического моделирования при анализе проблем естествознания (ПК-2);
- умение ориентироваться в современных алгоритмах компьютерной математики, совершенствовать, углублять и развивать математическую теорию, лежащую в их основе (ПК-7);
- собственное видение прикладного аспекта в строгих математических формулировках (ПК-8);
- способность к творческому применению, развитию и реализации математически сложных алгоритмов в современных программных комплексах (ПК-9);
- определение общих форм, закономерностей, инструментальных средств для групп дисциплин (ПК-10);
- владение методами математического и алгоритмического моделирования при анализе экономических и социальных процессов, задач бизнеса, финансовой и актуарной математики (ПК-11);

- умение формулировать в проблемно-задачной форме нематематические типы знания (в том числе гуманитарные) (ПК-14);
- возможность преподавания физико-математических дисциплин и информатики в общеобразовательных учреждениях, образовательных учреждениях начального профессионального, среднего профессионального и высшего профессионального образования на основе полученного фундаментального образования и научного мировоззрения (ПК-15);
- умение извлекать актуальную научно-техническую информацию из электронных библиотек, реферативных журналов (ПК-16).

**В результате освоения дисциплины обучающийся должен:**

- **Знать:**

направления вычислительной лингвистики, задачи автоматической обработки текста, системы информационного поиска и их архитектуру;

- **Уметь:**

оценивать системы информационного поиска с помощью тестовых коллекций, выполнять нормализацию текста;

находить, критически анализировать и объединять информацию из различных источников, ориентироваться в задачах автоматической обработки текста;

- **Владеть:**

навыками работы с современными вики-технологиями (Википедия, Викисловарь);

навыками представления знаний в проблемно-задачной форме, методами математического и алгоритмического моделирования

#### 4. Структура и содержание дисциплины «Интернет-математика»

Общая трудоемкость дисциплины составляет 5 зачетных единиц, 180 часов.

№ п/п	Раздел дисциплины	Семестр	Неделя Семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)	Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
1	Вычислительная лингвистика (CL). История развития лингвистики. Семиотика. Сравнительно-историческое языкознание. Разделы лингвистики. Смежные и пограничные науки. CL и обработка текста на естественном языке. Задачи и направления CL.	1	1	Лекция 1ч. Практика 1ч.	Учебная дискуссия

2	Информационный поиск (ИП). История ИП. 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.	1	2	Лекция 1ч. Практика 1ч. СРС – 4 часа	Обсуждение в аудитории Тест
3	Лауреаты премии Тьюринга.	1	3	Лекция 1ч. Практика 1ч. СРС – 24 часа	Доклады и презентации студентов
4	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.	1	4	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
5	Поисковые системы.	1	5	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
6	Вики-технологии: принципы работы, проблемы Википедии, сравнение с Semantic MediaWiki. Использование данных Википедии в компьютерной лингвистике.	1	6	Лекция 1ч. Практика 1ч. СРС – 14 часов	Обсуждение в аудитории Тест
7	Расширенные модели булева поиска: оператор близости терминов, проблемы булевых запросов. Метрики точности, полноты, fall-out.	1	7	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест

8	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	8	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
9	Оптимизация обработки булевых запросов. Пересечение списков словопозиций. Алгоритм слияния списков словопозиций.	1	9	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
10	Корпусная лингвистика: определение, классификация, типы, примеры корпусов.	1	10	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
11	Лексикография и корпуса. Толковые словари. Электронная лексикография.	1	11	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
12	Теория графов. Использование графов при решении задач автоматической обработки текстов.	1	12	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
13	Машинный перевод: стратегии, системы, перспективы.	1	14	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
14	Прикладные аспекты лексикографии.	1	15	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
15	Алгоритмы классификации полнотекстовых документов	1	16	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
16	Инструментальные системы разработки приложений по автоматической обработке текстов на естественном языке. Система обработки текстов GATE.	1	17	Лекция 1ч. Практика 1ч. СРС – 7 часов	Обсуждение в аудитории Тест
17	Семантический веб. Онтологическое моделирование. Работа в редакторе онтологий Protégé	1	18	Лекция 1ч. Практика 1ч. СРС – 20 часа	Итоговый тест
				<b>Всего 180 часов: Лекции – 17 ч., Практики – 17 ч., СРС – 146 ч.</b>	

## **5. Образовательные технологии**

- Традиционные и инновационные виды лекций
- Активные и интерактивные методы обучения: работа в группах, дебаты, полемика, учебная дискуссия.
- Балльно-рейтинговая система контроля.
- Доклады студентов, критика и обсуждение докладов
- Самостоятельная работа студентов (подготовка выступления по заданной теме, составление эссе, опорного конспекта, аннотации, реферата, тезисов, выполнение внеаудиторных упражнений, написание научно-популярной статьи в Википедии и т.п.)

## **6. Учебно-методическое обеспечение самостоятельной работы студентов. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины**

### **Самостоятельная работа:**

- Чтение специальной литературы.
- Проработка дополнительного материала по отдельным темам.
- Проработка пройденных лекционных материалов по конспекту лекций, учебникам и пособиям на основании вопросов, подготовленных преподавателем.
- Выполнение домашних контрольных работ.
- Подготовка группового отчета или презентации, написание эссе.
- Подготовка к промежуточному и итоговому контролю.
- Написание статьи в Википедии на основе учебников, пособий и научных статей.
- Рецензирование статей Википедии, написанных другими студентами.

### **Вопросы для подготовки к зачету**

1. Основные задачи и направления вычислительной лингвистики.
2. Дать определение и объяснить задачу информационного поиска.
3. Выполнить классификацию информационно-поисковых систем.
4. Перечислить шаги нормализации текста.
5. Архитектура информационно-поисковой системы.
6. Лауреаты премии Тьюринга (основные научные достижения, открытия).
7. Сравнить координатные и некоординатные инвертированные индексы.
8. Перечислить этапы построения инвертированного индекса.
9. Структуры данных для списков словопозиций: односвязные списки, списки с пропусками, массивы переменной длины.
10. Указать проблемы Википедии.
11. Сравнить принципы работы Википедии и Semantic MediaWiki.
12. Расширенные модели булева поиска и проблемы булевых запросов.
13. Оператор близости терминов.
14. Рассказать о вычислении метрик точности и полноты на примере.
15. Организация информации. Когда нужны и когда излишни внутренние ссылки.
16. Организация информации. Указать два типа интервик.
17. Организация информации. Указать типы ссылок между проектами Викимедиа.
18. Проект ВП «Связность»: определение тупиковых страниц и страниц-сирот, кластеры страниц.
19. Оптимизация обработки булевых запросов. Пересечение списков словопозиций.
20. Оптимизация обработки булевых запросов. Алгоритм слияния списков словопозиций.

21. Корпусная лингвистика: определение.
22. Корпусная лингвистика: классификация и типы корпусов.
23. Корпусная лингвистика: примеры корпусов.
24. Взаимосвязь лексикографии и корпусов.
25. Электронная лексикография.
26. Семантическое расстояние (на основе данных тезаурусов).
27. Использование графов при решении задач автоматической обработки текстов.
28. Машинный перевод: системы перевода, алгоритмы.
29. Алгоритмы классификации полнотекстовых документов.
30. Перспективы и проблемы семантический веба.

## **7. Учебно-методическое и информационное обеспечение дисциплины «Интернет-математика».**

### **Список литературы Основная**

1. Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А. Пескова О.В., Ягунова Е.В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие — М.: МИЭМ, 2011. — 272 с.

### **Дополнительная литература**

Электронно-библиотечная поисковая система «Академия Google».  
<http://scholar.google.com/>

1. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы: учеб. пособие для студентов линг.фак.вузов. —М.: Изд.центр «Академия», 2006. —304 с.
2. Маннинг К., Рагхаван П., Шютце Х. Введение в информационный поиск. — Вильямс, 2011. — ISBN 978-5-8459-1623-5.
3. Baeza-Yates R., Ribeiro-Neto B. Modern Information Retrieval // 1. Introduction. — Addison-Wesley, 1999. — ISBN 0-201-39829-X  
[ftp://mail.im.tku.edu.tw/seke/slide/baeza-yates/chap1\\_introductionmodern\\_ir.pdf](ftp://mail.im.tku.edu.tw/seke/slide/baeza-yates/chap1_introductionmodern_ir.pdf)
4. Леденёва В.В. Лексикография современного русского языка. Практикум: Учеб. пособие. - М.: Высшая школа, 2008. – 648 С.

Петрозаводский университет обеспечен необходимым комплектом лицензионного программного обеспечения.

### **Интернет-ресурсы**

1. Список тем по предмету «Интернет-математика» для самостоятельной работы студентов [http://ru.wikipedia.org/wiki/User:AKA\\_MBG/Todos](http://ru.wikipedia.org/wiki/User:AKA_MBG/Todos)
2. Список заданий по созданию и улучшению статей Википедии по темам предмета «Интернет-математика» для самостоятельной работы студентов [http://ru.wikipedia.org/wiki/User:AKA\\_MBG/task](http://ru.wikipedia.org/wiki/User:AKA_MBG/task)

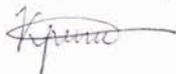
3. Презентации лекций по «Интернет-математике» для подготовки студентов.  
<http://mathem.krc.karelia.ru/section.php?plang=r&id=925>
4. Группа «Интернет-математика в ПетрГУ» в социальной сети «ВКонтакте» для взаимодействия преподавателя и студентов <https://vk.com/club41102811>
5. Презентации студентов по выбранным темам по предмету «Интернет-математика»  
[http://whinger.narod.ru/paper/cl/slides\\_wp\\_topics.doc.html](http://whinger.narod.ru/paper/cl/slides_wp_topics.doc.html)
6. Презентации студентов по лауреатам премии Тьюринга  
[http://whinger.narod.ru/paper/cl/Turing\\_Award\\_presentations.doc.html](http://whinger.narod.ru/paper/cl/Turing_Award_presentations.doc.html)

#### **8. Материально-техническое обеспечение дисциплины**

Петрозаводский университет обеспечен необходимыми материально-техническими ресурсами: мультимедиапроектор, ноутбук для проведения лекций и докладов студентов в виде презентаций в программе Power Point, электронные диски, флеш карты, экран. Научная библиотека ПетрГУ располагает обширным фондом учебной и научной литературы по курсу.

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего профессионального образования (ФГОС ВПО) направления **010100 Математика** (квалификация Магистр), 2010 г. с учетом методических рекомендаций и Примерной основной образовательной программы ВПО по направлению **010100 Математика** (квалификация Магистр).

Автор: Крижановский Андрей Анатольевич, старший преподаватель, кандидат технических наук.



Программа рассмотрена и утверждена на заседании кафедры теории вероятностей и анализа данных

«30» мая 2012 года, протокол № 5.

Зав. кафедрой



Рогов А.А.

Программа одобрена на заседании учебно-методической комиссии математического факультета «25» июня 2012 года, протокол № 10.

Председатель учебно-методической комиссии математического факультета,

доцент, к.ф.-м.н. Семенова Е.Е.

