**Санкт-Петербургский государственный университет**

**Р А Б О Ч А Я П Р О Г Р А М М А**

**УЧЕБНОЙ ДИСЦИПЛИНЫ**

Методы информационного поиска

Information Retrieval

**Язык(и) обучения**

русский

Трудоемкость в зачетных единицах: 4

Регистрационный номер рабочей программы: 027265

Санкт-Петербург

2020

**Раздел 1. Характеристики учебных занятий**

**1.1. Цели и задачи учебных занятий**

Дисциплина «Методы информационного поиска» представляет обучающимся комплекс знаний, умений и навыков, позволяющих овладеть основами методов ведения информационного поиска. По результатам данного курса обучающиеся смогут научиться выбирать модели, адекватные для решения различных задач, и применять эти знания на практике. Такие знания необходимы для успешной разработки высоконадежных информационных систем.

Отдельные параметры курса могут варьироваться по степени сложности в зависимости от уровня подготовки обучающихся.

Курс должен быть построен на принципах компетентного, деятельного подхода к реализации документных информационных систем.

**1.2. Требования подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)**

Программа курса предназначена обучающимся магистратуры и рассчитана на лиц, изучавших основы разработки программного обеспечения в объеме четырех курсов бакалавриата и владеющих базовыми навыками работы с компьютером, а также теоретическими основами и практическими навыками применения систем управления базами данных.

Максимальная эффективность программы будет обеспечена при условии, что обучающийся:

- владеет основами теории графов;

- владеет методами теории алгоритмов;

- владеет основами теории баз данных.

**1.3. Перечень результатов обучения (learning outcomes)**

**1.3.1. Компетенции, развиваемые дисциплиной**

Компетенции, которые формируются у обучающихся в процессе изучения дисциплины:

* ПКА-1 Способен находить, формулировать и решать актуальные проблемы фундаментальной и прикладной информатики и информационных технологий
* ПКП-1 Способен демонстрировать фундаментальные знания математических и естественных наук, программирования и информационных технологий
* ПКП-8 Способен использовать основные концептуальные положения функционального, логического, объектно-ориентированного и визуального направлений программирования, методы, способы и средства разработки программ в рамках этих направлений

**1.3.2. Знания, умения и навыки, формируемые дисциплиной**

Понимание идей и этапов построения поисковой системы: обработка текстов, построение индекса, поиск по индексу. Приобретение навыков реализации простейшей системы информационного поиска.

**1.4. Перечень и объём активных и интерактивных форм учебных занятий**

Активные формы учебных занятий — лекции, предполагающие обсуждение материала с преподавателем, 10 ак.ч.

**Раздел 2. Организация, структура и содержание учебных занятий**

**2.1. Организация учебных занятий**

**2.1.1 Основной курс**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся | | | | | | | | | | | | | | | | | | |
| Код модуля в составе дисциплины,  практики и т. п. | Контактная работа обучающихся с преподавателем | | | | | | | | | | Самостоятельная работа | | | | | | Объём активных и интерактивных  форм учебных занятий | Трудоёмкость |
| лекции | семинары | консультации | Практические занятия | лабораторные работы | контрольные работы | коллоквиумы | текущий контроль | промежуточная  аттестация | итоговая аттестация | под руководством преподавателя | в присутствии  преподавателя | сам. раб. с использованием  методических материалов | текущий контроль (сам. раб.) | промежуточная аттестация (сам. раб.) | итоговая аттестация  (сам. раб.) |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | | | | | | | | | | | | | |
| Форма обучения: очная | | | | | | | | | | | | | | | | | | |
| Семестр 2 | 30 |  | 2 | 15 |  |  |  |  | 2 |  |  |  | 65 |  | 30 |  | 10 | 4 |
|  | 1-100 |  | 1-100 | 1-30 |  |  |  |  | 1-100 |  |  |  | 1-1 |  | 1-1 |  |  |  |
| ИТОГО | 30 |  | 2 | 15 |  |  |  |  | 2 |  |  |  | 65 |  | 30 |  | 10 | 4 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Виды, формы и сроки текущего контроля успеваемости и промежуточной аттестации | | | | | | |
| Код модуля в составе дисциплины, практики и т.п. | Формы текущего контроля успеваемости | | Виды промежуточной аттестации | | Виды итоговой аттестации  (только для программ итоговой аттестации и дополнительных образовательных программ) | |
| Формы | Сроки | Виды | Сроки | Виды | Сроки |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | |
| Форма обучения: очная | | | | | | |
| Семестр 2 |  |  | экзамен, устно, традиционная форма | по графику промежуточной аттестации |  |  |

**2.2. Структура и содержание учебных занятий**

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Наименование темы (раздела, части) | Вид учебных занятий | Количество часов |
| I | Булев поиск | лекции | 8 |
| практические занятия | 4 |
| по методическим материалам | 16 |
| II | Обработка текстов | лекции | 8 |
| практические занятия | 4 |
| по методическим материалам | 16 |
| III | Машинное обучение | лекции | 8 |
| практические занятия | 4 |
| по методическим материалам | 16 |
| IV | Семантическое индексирование | лекции | 6 |
| практические занятия | 3 |
| по методическим материалам | 17 |
|  | Промежуточная аттестация | самостоятельная работа | 30 |
| консультации | 2 |
| экзамен | 2 |
| **Итого** | | | **144** |

**2.2.1 Содержание учебных занятий**

Темы для изучения и обсуждения

Раздел 1. Булев поиск.

* Булев поиск.
* Индекс – построение и сжатие.

Раздел 2. Обработка текстов.

* Лексикон, словари, нечеткий поиск
* Векторная модель информационного поиска
* Ранжирование

Раздел 3. Машинное обучение.

* Вероятностная модель информационного поиска
* Классификация текстов
* Кластеризация текстов
* Метод опорных векторов и машинное обучение на документах

Раздел 4. Семантическое индексирование.

* Латентно-семантическое индексирование
* Анализ ссылок

**Раздел 3. Обеспечение учебных занятий**

**3.1. Методическое обеспечение**

**3.1.1 Методические указания по освоению дисциплины**

Успешное освоение дисциплины возможно благодаря посещению лекций и практических занятий, участию в обсуждении рассматриваемых вопросов, самостоятельной работе, включающей в себя чтение специальной литературы по разделам темы.

**3.1.2 Методическое обеспечение самостоятельной работы**

Самостоятельная работа обучающихся, как вид деятельности, стимулирующий активность, самостоятельность, познавательный интерес с целью поиска необходимой информации, приобретения знаний, использования этих знаний для решения учебных, научных и профессиональных задач, представляет собой важную составляющую учебного процесса, которой отводится не менее половины учебного времени при очной форме обучения. Время, отводимое на самостоятельную работу, должно использоваться обучающимися для наиболее полного освоения учебной дисциплины. Следовательно, организация эффективной внеаудиторной самостоятельной работы в процессе обучения требует создания условий, призванных обеспечить рациональное и планомерное управление учебной деятельностью, протекающей в отсутствие преподавателя.

Роль преподавателя в организации самостоятельной работы состоит в координации действий обучающихся в освоении дисциплины, в методическом и организационном обеспечении учебного процесса. Взаимодействие между преподавателем и обучающимся осуществляется в форме консультаций, а также через сайт поддержки курса. Преподаватели также оказывают помощь обучающимся по планированию и организации самостоятельной работы.

**3.1.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания**

**3.1.3.1. Методика проведения текущего контроля успеваемости и промежуточной аттестации**

*Методика проведения контроля успеваемости практической части курса*

В течение семестра обучающийся выполняет два практических задания на тему информационного поиска. Для этого он выбирает язык программирования, согласовывает его с преподавателем и к установленному сроку демонстрирует работоспособность своей реализации.

*Методика проведения контроля качества докладов*

В течение семестра обучающийся должен сделать два доклада на актуальную тему в предметной области. Для этого он самостоятельно выбирает относительно свежую статью из топовых тематических конференций, согласовывает свой выбор с преподавателем и делает по ней доклад в установленное время длительностью не менее 20 минут и не более одного академического часа.

**3.1.3.2. Критерии оценивания итогового процента освоения дисциплины**

За каждое успешно выполненное практическое задание выставляется базовая оценка 7 баллов. При использовании современных средств и методов программирования оценка может быть повышена до 10 баллов.

Каждый доклад оценивается от 0 до 5 баллов в зависимости от продемонстрированного понимания материала.

В случае нарушения сроков сдачи задания оценка может быть снижена на усмотрение преподавателя.

Итоговая оценка рассчитывается из суммы всех полученных за программную реализацию и доклад баллов согласно таблице:

|  |  |  |
| --- | --- | --- |
| Итоговый процент  выполнения, % | Оценка СПбГУ при  проведении экзамена | Оценка  ECTS |
| 90-100 | отлично | A |
| 80-89 | хорошо | B |
| 70-79 | хорошо | C |
| 61-69 | удовлетворительно | D |
| 50-60 | удовлетворительно | E |
| менее 50 | неудовлетворительно | F |

**3.1.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)**

**3.1.4.1. Формируемые дисциплиной компетенции**

**ПКА-1** Способен находить, формулировать и решать актуальные проблемы фундаментальной и прикладной информатики и информационных технологий

□ Формируется дисциплиной.

✓ Развивается дисциплиной.

□ Полностью сформирована по результатам освоения дисциплины.

Шкала оценивания: линейная, определяется долей успешно выполненных заданий, проверяющих данную компетенцию.

**ПКП-1** Способен демонстрировать фундаментальные знания математических и естественных наук, программирования и информационных технологий

□ Формируется дисциплиной.

✓ Развивается дисциплиной.

□ Полностью сформирована по результатам освоения дисциплины.

Шкала оценивания: линейная, определяется долей успешно выполненных заданий, проверяющих данную компетенцию.

**ПКП-8** Способен использовать основные концептуальные положения функционального, логического, объектно-ориентированного и визуального направлений программирования, методы, способы и средства разработки программ в рамках этих направлений

□ Формируется дисциплиной.

✓ Развивается дисциплиной.

□ Полностью сформирована по результатам освоения дисциплины.

Шкала оценивания: линейная, определяется долей успешно выполненных заданий, проверяющих данную компетенцию.

**3.1.4.2. Контрольно-измерительные материалы (примеры)**

*Примеры заданий*

* «Реализация системы, выполняющей булев поиск по заданному корпусу документов».
* «Реализация ранжирования с помощью методов машинного обучения».

*Примеры статей для разбора*

* Lofgren P. et al, Personalized PageRank Estimation and Search: A Bidirectional Approach, WSDM-16
* Johnson J. et al, Billion-scale similarity search with GPUs, SIGIR-2017

**3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса**

Для оценки обучающимися содержания и качества учебного процесса может применяться анкетирование в соответствии с методикой и графиком, утвержденными в установленном порядке.

**3.2. Кадровое обеспечение**

**3.2.1 Образование и (или) квалификация штатных преподавателей и иных лиц, допущенных к проведению учебных занятий**

К преподаванию дисциплины могут быть допущены преподаватели, имеющие диплом о высшем образовании по соответствующему направлению.

**3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом**

Специальных требований нет.

**3.3. Материально-техническое обеспечение**

**3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий**

В аудиториях, где проводятся занятия, необходимо наличие досок и средств письма на них.

**3.3.2 Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования**

Аудитории для проведения занятий должны быть оснащены проекционной техникой и компьютером и возможности подключения дополнительного портативного компьютера и возможности беспроводного доступа в сеть интернет.

**3.3.3 Характеристики специализированного оборудования**

Специальных требований нет

**3.3.4 Характеристики специализированного программного обеспечения**

Специальных требований нет

**3.3.5 Перечень и объёмы требуемых расходных материалов**

Мел, маркеры для досок, губка

**3.4. Информационное обеспечение**

**3.4.1 Список обязательной литературы**

1. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.
2. David A. Grossman, Ophir Frieder, Information Retrieval: Algorithms and Heuristics, Springer, 1998

**3.4.2 Список дополнительной литературы**

1. G.G. Chowdhury, Introduction to Modern Information Retrieval, Third Edition Paperback, 2010

**3.4.3 Перечень иных информационных источников**

1. http://forum.searchengines.ru/archive/index.php/t-77275.html.

**Раздел 4. Разработчики программы**

Михайлова Елена Георгиевна, к.ф.-м.н., доцент кафедры информационно-аналитических систем СПбГУ, [e.mikhaylova@spbu.ru](mailto:e.mikhaylova@spbu.ru). +7 (921) 9979791.