**Санкт-Петербургский государственный университет**

**Р А Б О Ч А Я П Р О Г Р А М М А**

**УЧЕБНОЙ ДИСЦИПЛИНЫ**

Методы хранения и индексирования

Data Structures and Indexing

**Язык(и) обучения**

русский

Трудоемкость в зачетных единицах: 3

Регистрационный номер рабочей программы: 040705

Санкт-Петербург

2020

**Раздел 1. Характеристики учебных занятий**

**1.1. Цели и задачи учебных занятий**

Cформировать у слушателей представление о методах хранения информации, применения различных методов индексирования для организации быстрого доступа к данным.

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

Знание основ информатики, программирования, курса «Базы данных и СУБД» (или аналогичного) в пределах бакалаврской подготовки.

Дисциплина «Методы хранения и индексирования» («МХИ») является семинаром по выбору в подготовке профессионального математика-программиста.  
Наилучший результат может быть достигнут, если обучающимся изучен курс «Разработка приложений СУБД».

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

В процессе изучения дисциплины «Методы хранения и индексирования» обучаемые приобретают следующие

***знания***

* основные методы получения, хранения и переработки информации;
* основные принципы построения индексов;
* применение различных видов индексирования.

***умения***

* использовать средства современных СУБД для организации хранения различных видов информации;
* анализировать необходимость применения различных методов индексирования для ускорения доступа к данным;
* создавать и реализовывать новые схемы индексирования и применять их к данным;
* делать анализ и грамотную оценку эффективности применения индексов.

***навыки***

* применения средств современных СУБД для эффективной работы с данными;
* осуществления целенаправленного поиска информации о технологических достижениях в сети Интернет и из других источников;
* взаимодействия с коллегами, работы в коллективе.

Знать содержание дисциплины «Методы хранения и индексирования», в частности, знать различные способы представления данных средствами современных СУБД, применять современные технологии для организации доступа к данным, иметь возможность организовывать новые эффективные структуры для ускорения доступа к данным.

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

В качестве основных интерактивных форм (30 ак. часов) предполагается проведение семинарских занятий, на которых обучающиеся будут делать доклады о разных методах индексирования на основании прочитанных статей из научных журналов и трудов конференций по БД, будут применять различные методы индексирования на основе современных СУБД.

**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 7 |  | 30 |  |  |  |  |  |  | 2 |  |  |  | 73 |  | 3 |  | 30 | 3 |
|  |  | 2-25 |  |  |  |  |  |  | 1-25 |  |  |  | 1-1 |  | 1-1 |  |  |  |
| ИТОГО |  | 30 |  |  |  |  |  |  | 2 |  |  |  | 73 |  | 3 |  |  | 3 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 7 |  |  | зачёт, устно, традиционная форма | по графику промежуточной аттестации |  |  |

**2.2. Структура и содержание учебных занятий**

Применение различных типов данных для хранения информации на основе СУБД

* Организация хранения при помощи средств современных СУБД иерархически организованных данных, географических данных, текстовых данных большого объема.

1. Применение одномерных способов индексирования

* Построение индексов на основе B-деревьев, битовых шкал, хеширования. Сравнение эффективности рассматриваемых технологий.

1. Применение пространственных способов индексирования

* Построение пространственных индексов на основе R-дерева и Spatial-grid.

1. Способы многомерного индексирования данных

* Построение структур для многоключевого поиска информации. Метрические индексы.

1. Применение полнотекстового индексирования

* Построение и настройка полнотекстовых индексов, ведение словаря для поиска текстовой информации.

**Раздел 3. Обеспечение учебных занятий**

**3.1. Методическое обеспечение**

**3.1.1 Методические указания по освоению дисциплины**

Успешное освоение дисциплины возможно благодаря посещению семинаров, участию в обсуждении рассматриваемых вопросов, самостоятельной работе, включающей в себя чтение специальной литературы по разделам темы, выполнению докладов на заданную тему, применению изученных навыков на практических занятиях.

**3.1.2 Методическое обеспечение самостоятельной работы**

Самостоятельная работа обучающихся в рамках данной дисциплины является важным компонентом обучения, предусмотренным компетентностно-ориентированным учебным планом и рабочей программой учебной дисциплины.

Настоящей программой предусмотрены формы самостоятельной работы с использованием методических материалов по тематике курса и источников, указанных в обязательной, дополнительной литературе и интернет-источниках, указанных с данной программе.

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

Для получения зачета учащийся должен сделать не менее двух докладов на семинарах и построить изученные индексные структуры на основе реляционных СУБД.

За каждый доклад слушатель может получить максимум 20 баллов. Оценка может быть снижена за не полностью описанные возможности индексов в разных СУБД, не указаны все параметры, отсутствие сравнения производительности. Итого 40 баллов максимум за два доклада.

За реализацию каждого метода индексирования слушатель получает не более 15 баллов.

Общая оценка складывается из оценки за доклады и за практическую реализацию различных методов индексирования.

|  |  |  |
| --- | --- | --- |
| Итоговый процент  выполнения, % | Оценка СПбГУ при  проведении зачёта | Оценка  ECTS |
| 90-100 | зачтено | A |
| 80-89 | зачтено | B |
| 70-79 | зачтено | C |
| 61-69 | зачтено | D |
| 50-60 | зачтено | E |
| менее 50 | не зачтено | F |

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

Доклады по темам:

* Организация хранения при помощи средств современных СУБД иерархически организованных данных, географических данных, текстовых данных большого объема.
* Применение одномерных способов индексирования
* Построение индексов на основе B-деревьев, битовых шкал, хеширования. Сравнение эффективности рассматриваемых технологий.
* Применение пространственных способов индексирования
* Построение пространственных индексов на основе R-дерева и Spatial-grid.
* Способы многомерного индексирования данных
* Построение структур для многоключевого поиска информации. Метрические индексы.
* Применение полнотекстового индексирования
* Построение и настройка полнотекстовых индексов, ведение словаря для поиска текстовой информации

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

Для оценки обучающимися содержания и качества учебного процесса применяется 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**

Для проведения занятий требуется аудитория, оснащенная проектором и компьютерами.

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

MS Office.

**3.3.3 Характеристики специализированного оборудования**

Необходимо наличие компьютерных классов с высокопроизводительными многопроцессорными компьютерами/параллельными кластерами.

**3.3.4 Характеристики специализированного программного обеспечения**

На компьютерах должны быть установлены СУБД Microsoft SQL Server, Express Edition, Postgres.

**3.3.5 Перечень и объёмы требуемых расходных материалов**

Не требуются.

**3.4. Информационное обеспечение**

**3.4.1 Список обязательной литературы**

1. Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце. Введение в информационный поиск. Вильямс. 2010. Пер. с англ. Д. Клюшина.

**3.4.2 Список дополнительной литературы**

1. Gordon S. Linoff, Michael J. A. Berry -Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 3rd Edition. Wiley Publishing, Inc, 2011

**3.4.3 Перечень иных информационных источников**

1. <http://www.vldb.org/> — сайт конференции Very large data bases.
2. <http://syrcodis.ispras.ru/> — сайт Spring Colloquium for Young Researchers on Database and Information Systems.
3. <http://ieee-icde2014.eecs.northwestern.edu/index.html> — сайт International Conference on Data Engineering.
4. <http://www.edbt.org/> — Extending Database Technology.

**Раздел 4. Разработчики программы**

Михайлова Елена Георгиевна, доцент кафедры ИАС СПбГУ e.mikhaylova@spbu.ru, тел. +7921-9979791.