**Санкт-Петербургский государственный университет**

**Р А Б О Ч А Я П Р О Г Р А М М А**

**УЧЕБНОЙ ДИСЦИПЛИНЫ**

Методы организации информации

Information Management Methods

**Язык(и) обучения**

русский

Трудоемкость в зачетных единицах: 4

Регистрационный номер рабочей программы: 027257

Санкт-Петербург

2020

**Раздел 1. Характеристики учебных занятий**

**1.1. Цели и задачи учебных занятий**

Дисциплина «Методы организации информации» представляет обучающимся комплекс знаний, умений и навыков, позволяющих овладеть фундаментальными основами методов логического анализа и аналитической обработки данных в информационных системах и системах обработки «больших данных». Такие знания необходимы для успешной разработки высокоэффективных систем аналитической обработки больших данных

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 высокоэффективных информационно-аналитических систем.

Для достижения поставленных целей в условиях ограниченных ресурсов, что предполагает распределение содержания обучения по следующим видам деятельности: изучение основных концепций организации информации, изучение специфических вопросов логического анализа данных, выполнение индивидуальных заданий, выполнение коллективных заданий (вне аудитории и в аудитории).

Основным методологическим принципом построения программы данного курса, равно как и всей концепции обучения, является принцип поэтапного системного накопления знаний и формирования необходимых компетенций по модели: от простого и/или знакомого – к сложному и/или незнакомому, а основной методологической стратегией прохождения отдельных разделов программы является ступенчатость и цикличность, предусматривающие постепенный возврат к ранее усвоенному материалу на более высоком концептуальном уровне.

Цель изучения дисциплины: обучение теоретическим основам методов организации информации, методов логического анализа данных и информационных системах и методам организации хранения и поиска больших данных, а также приобретение навыков практического использования информационно-аналитических систем.

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

Программа курса предназначена обучающимся магистратуры и рассчитана на обучающихся, 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 методов статистической обработки данных;
* владеет методами теории алгоритмов и знаком с основными структурами хранения данных.

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

Основной целью курса является овладение обучающимися теоретическими основаниями методов организации информации и логического анализа данных и получение навыков применения полученных знаний, а именно:

* знание содержания дисциплины «методы организации информации» и возможностей применения отдельных разделов курса;
* навыки применения информационно-аналитических систем;
* умение выполнения основных задач анализа данных на практике;
* развитие способности дальнейшего освоения методов организации информации.

Курс способствует формированию следующих компетенций:

* ОПК-1 – способен находить, формулировать и решать актуальные проблемы фундаментальной и прикладной информатики и информационных технологий;
* 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-7 – способен использовать знания направлений развития компьютеров с традиционной (нетрадиционной) архитектурой; современных системных программных средств: операционных систем, операционных и сетевых оболочек, сервисных программ; тенденции развития функций и архитектур проблемно-ориентированных программных систем и комплексов в профессиональной деятельности;
* 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.

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

Занятия в активных формах: лекции, предполагающие дискуссию с преподавателем, 10 академических часов.

**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 1 | 30 |  | 2 | 15 |  |  |  |  | 2 |  |  |  | 58 |  | 37 |  | 10 | 4 |
|  | 2-100 |  | 2-100 | 10-25 |  |  |  |  | 2-100 |  |  |  | 1-1 |  | 1-1 |  |  |  |
| ИТОГО | 30 |  | 2 | 15 |  |  |  |  | 2 |  |  |  | 58 |  | 37 |  |  | 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 1 |  |  | экзамен, устно, традиционная форма | по графику промежуточной аттестации |  |  |

**2.2. Структура и содержание учебных занятий**

Модуль 1

1. Базовые алгоритмы кластеризации на основе центров кластеров

2. Плотностные алгоритмы кластеризации

3. Кластеризация в пространствах высокой размерности

Модуль 2

1. Методы и алгоритмы классификации

2. Применение методов классификации в задачах обработки текстов

3. Методы анализа потоков данных

Модуль 3

1. Структуры хранения в базах данных для аналитической обработки

2. Методы индексирования для пространств большой размерности

3. Методы поиска ближайших объектов

4. Выполнение запросов на основе подобия объектов

Модуль 4

1. Извлечение и очистка данных в распределенных системах

2. Методы извлечения, преобразования и загрузки данных

3. Обработка слабоструктурированной информации.

**Раздел 3. Обеспечение учебных занятий**

**3.1. Методическое обеспечение**

**3.1.1 Методические указания по освоению дисциплины**

Методические материалы включают в себя следующие типы материалов — литература, Интернет-ресурсы, учебные пособия, с опорой на которые проводится аудиторная работа.

**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.

К числу методических пособий относятся:

* задания для самостоятельного выполнения;
* литература по теме курса;
* сайт поддержки курса в сети интернет.

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**

Итоговый процент освоения дисциплины складывается из трёх факторов: оценки за домашние задания, оценки за текущий контроль и оценки за теоретический экзамен. Домашние задания оцениваются из шкал оценивания, указанных для каждой задачи отдельно. Критерии оценивания и максимальные баллы за каждую задачу доводятся до обучающихся вместе с условием задачи. Итоговый балл за домашние задания вычисляется как сумма баллов за все задания. Для командных заданий всем членам команды ставится одинаковая оценка.

Оценка за текущий контроль ставится по шкале от 0 до 10 по критериям, приведённым в разделе 3.1.4. Всем членам команды ставится одинаковая оценка.

Оценка за экзамен ставится по следующим правилам: ответ на каждый вопрос билета и на дополнительные вопросы оценивается по шкале от 0 (нет ответа) до 10 (очень хороший ответ), далее оценка усредняется. Результат переводится в диапазон от 0 до 100.

Итоговый процент выполнения дисциплины вычисляется как взвешенная сумма оценок за домашние задания, текущий контроль и экзамен по следующей формуле: итоговая оценка = min(100, 0.3 \* (оценка за домашние задания) + 0.1 \* (оценка за текущий контроль) + 0.8 \* (оценка за экзамен)). Перед вычислением итогового процента освоения оценка за домашние задания, оценка за текущий контроль и оценка за экзамен приводятся к диапазону 0–100 линейным преобразованием. Далее применяется следующее правило выставления оценки:

|  |  |  |
| --- | --- | --- |
| **Итоговый процент выполнения, %** | **Оценка СПбГУ при проведении экзамена** | **Оценка ECTS** |
| 90-100 | отлично | А |
| 80-89 | хорошо | В |
| 70-79 | хорошо | С |
| 61-69 | удовлетворительно | D |
| 50-60 | удовлетворительно | Е |
| менее 50 | неудовлетворительно | F |

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

Формируемые дисциплиной компетенции:

ОПК-1 – способен находить, формулировать и решать актуальные проблемы фундаментальной и прикладной информатики и информационных технологий;

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-7 – способен использовать знания направлений развития компьютеров с традиционной (нетрадиционной) архитектурой; современных системных программных средств: операционных систем, операционных и сетевых оболочек, сервисных программ; тенденции развития функций и архитектур проблемно-ориентированных программных систем и комплексов в профессиональной деятельности;

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.

Шкалы оценивания всех компетенций линейные, степень сформированности определяется долей успешно выполненных заданий, проверяющих данную компетенцию.

*Перечень примерных вопросов для экзамена:*

1. Основные понятия – информация, данные, знания. Виды информации. Обработка данных и ее виды. Data Mining. Классификация задач Data Mining.
2. Модели процессов обработки данных. Модель: конечные автоматы.
3. Модели процессов обработки данных. Модель: сети Петри.
4. Задачи обработки данных различных типов. Прикладные области обработки данных. Оцифровка сигналов. Теорема Котельникова.
5. Базы данных. OLTP – системы. Неэффективность OLTP для анализа данных. Определение и свойства хранилищ данных.
6. Физические и виртуальные хранилища данных (ХД). Основные проблемы создания ХД. Витрины данных.
7. Данные в хранилищах данных. ETL процесс.
8. Представление данных в виде гиперкуба. Операции над гиперкубом. Пример. Технология OLAP. Тест FASMI.
9. Многомерное представление данных и многомерный куб. Представление данных в виде гиперкуба. Пример.
10. Основные понятия гиперкубов (OLAP кубов). Структура OLAP куба. Операции над гиперкубом.
11. Архитектура OLAP. Компоненты OLAP. MOLAP, ROLAP, HOLAP.
12. Постановка задачи классификации и регрессии. Алгоритм построения 1-правил.
13. Постановка задачи классификации и регрессии. Метод Naive-Base.
14. Постановка задачи классификации и регрессии. Деревья решений. Метод разделяй и властвуй.
15. Постановка задачи классификации и регрессии. Деревья решений. Метод ID3.
16. Постановка задачи классификации и регрессии. Деревья решений. Алгоритм покрытия.
17. Постановка задачи классификации и регрессии. Прогнозирование временных рядов.
18. Постановка задачи поиска ассоциативных правил. Сиквенциальный анализ. Разновидности задачи поиска ассоциативных правил.
19. Постановка задачи поиска ассоциативных правил. Представление результатов.
20. Постановка задачи поиска ассоциативных правил. Алгоритм Apriori.
21. Постановка задачи кластеризации. Меры близости. Представление результатов.
22. Постановка задачи кластеризации. Алгоритмы кластеризации: иерархические и неиерархические.
23. Постановка задачи кластеризации. Алгоритм k-means.
24. Постановка задачи кластеризации. Алгоритм Fuzzy C-Means.
25. Задача анализа текстов. Этапы анализа. Предобработка текста.
26. Извлечение ключевых понятий из текста.
27. Классификация текстовых документов. Методы классификации текстовых документов.
28. Средства синхронизации потоков в .NET и С#.
29. Распределенная и параллельная обработка запросов, распределенная обработка транзакций.

*Проверяемые компетенции*: ОПК-1, ПКА-1, ПКП-1, ПКП-7, ПКП-8

*Критерии оценивания*: совпадают с критериями оценивания ответа на экзамене.

**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**

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

**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**

Требуется возможность использования систем управления базами данных (Oracle, MS SQL Server, Postgres SQL).

**3.3.5 Перечень и объёмы требуемых расходных материалов**

Специальных требований нет

**3.4. Информационное обеспечение**

**3.4.1 Список обязательной литературы**

Не требуется.

**3.4.2 Список дополнительной литературы**

Не требуется.

**3.4.3 Перечень иных информационных источников**

Не требуется.

**Раздел 4. Разработчики программы**

Серов Михаил Александрович, к.т.н., доцент, кафедра системного программирования СПбГУ, m.serov@spbu.ru, +7 (962) 2951889