

**РАБОЧАЯ ПРОГРАММА
УЧЕБНОЙ ДИСЦИПЛИНЫ**

*Алгоритмы и технологии анализа данных
Algorithms and Technologies of Data Analysis*

Язык(и) обучения

русский

Трудоёмкость (границы трудоёмкости) в зачетных единицах: ____2____

Регистрационный номер рабочей программы: 001130

Санкт-Петербург

2016

Раздел 1. Характеристики учебных занятий

1.1. Цели и задачи учебных занятий

Изучение основ алгоритмов и технологий для анализа коллекций с данными. Освоение алгоритмов интеллектуального анализа текстовых данных. Командная разработка различных методов и технологий поиска скрытых данных в коллекциях документов больших объемов. Приобретение навыков практического применения изученных алгоритмов.

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

Освоение дисциплин: Теория вероятности, Линейная и компьютерная алгебра, Дискретная математика, Программирование.

1.3. Перечень результатов обучения (learning outcomes)

Дисциплина способствует формированию следующих компетенций:

- ОКМ-4 Готов самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях;
- ПК-1 способность применять в профессиональной деятельности современные языки программирования и языки баз данных, методологии системной инженерии, системы автоматизации проектирования, электронные библиотеки и коллекции, сетевые технологии, библиотеки и пакеты программ, современные профессиональные стандарты информационных технологий в соответствии с профилем подготовки;
- ПК-5 способность использовать углубленные теоретические и практические знания в области информационных технологий и прикладной математике, фундаментальные концепции и системные методологии, международные и профессиональные стандарты в области информационных технологий, а также знания, которые находятся на передовом рубеже данной науки.

После изучения курса обучающиеся получают:

- знания известных алгоритмов и технологий анализа данных,
- умения применять на практике изученные алгоритмы,
- навыки командной разработки технологий анализа данных.

1.4. Перечень активных и интерактивных форм учебных занятий

Все семинары и практические занятия необходимо проводить с привлечением интерактивных методов: групповое обсуждение докладов, работа в малых группах, представление самостоятельно выполненных индивидуальных заданий и коллективное обсуждение полученных результатов.

Практические занятия, 16 часов.

Семинары, 12 часов.

Раздел 2. Организация, структура и содержание учебных занятий

2.1. Организация учебных занятий

Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся																		
Период обучения (модуль)	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																		
очная форма обучения																		
Семестр 2		12		16				4	2				28	4	6		16	2
ИТОГО		12		16				4	2				28	4	6		16	2

Формы текущего контроля успеваемости, виды промежуточной и итоговой аттестации			
Период обучения (модуль)	Формы текущего контроля успеваемости	Виды промежуточной аттестации	Виды итоговой аттестации (только для программ итоговой аттестации и дополнительных образовательных программ)
ОСНОВНАЯ ТРАЕКТОРИЯ			
очная форма обучения			
Семестр 2	Командная презентация промежуточных результатов и разработанного комплекса программ по проектам	зачет	

2.2. Структура и содержание учебных занятий

Период обучения (модуль): **Семестр 2**

№ п/п	Наименование темы (раздела, части)	Вид учебных занятий	Количество часов
1	Первичное знакомство с алгоритмами и технологиями анализа текстовых данных. Парное программирование.	семинары	6
		практические занятия	8

	<p>Семинары:</p> <p>1.1 Понятие документа, модели документа. Текстовые характеристики документа. Модель документа на множестве слов. Проблемы выделения слов в документе. Модель, основанная на стемминге документа, алгоритм Портера стемминг.</p> <p>1.2 Модель на взвешивании слов документа. Алгоритм TF - IDF и его разновидности. Алгоритмы поиска устоявшихся словосочетаний в текстовых документах.</p> <p>1.3 Задачи кластерного анализа текстовых коллекций больших объемов. Теоретико-вероятностные методы кластерного анализа данных (алгоритм K-средних, алгоритм K-medians, EM-алгоритм).</p> <p>1.4 Модель понимания информации Shu-Ha-Ri. Парное программирование, командная разработка ПО.</p> <p>Практические занятия:</p> <p>1.5 Практика парного программирования.</p> <p>1.6 Анализ и обработка текстовых данных. Реализация модели документа на множестве слов.</p> <p>1.7 Командное создание и презентация продукта.</p> <p>1.8 Алгоритм взвешивания слов TF - IDF с учетом их стемминга.</p> <p>1.9 Использование классического метода K-средних и EM-алгоритма для кластерного анализа данных. Оценка качества реализованных методов кластерного анализа данных на коллекции с объемом документов более 10000.</p>	по методическим материалам	14
2	<p>Алгоритмы и технологии анализа данных на основе систем искусственного интеллекта. Экстремальное программирование.</p> <p>Семинары:</p> <p>1.1 Кластерный анализ текстовых данных с использованием технологий на основе Нейронных сетей.</p> <p>1.2 Кластерный анализ текстовых данных с использованием генетических алгоритмов.</p> <p>1.3 Классические и гибкие методологии разработки ПО. Оценка необходимого размера команды. Основные роли в разработке ПО. Экстремальное программирование.</p> <p>Практические занятия:</p> <p>1.4 Использование нейронной сети Кохонена и генетического алгоритма для кластерного анализа текстовых коллекций документов.</p> <p>1.5 Оценки качества реализованных методов</p>	семинары	6
		практические занятия	8
		по методическим материалам	14

	кластеризации и их сравнение. 1.6 Тестирование производительности кластерного анализа данных на малых, средних и больших объемах тестовых коллекций. 1.7 Представление командной методологии разработки. 1.8 Unit-тестирование разработанного программного комплекса.		
--	--	--	--

Раздел 3. Обеспечение учебных занятий

3.1. Методическое обеспечение

3.1.1 Методические указания по освоению дисциплины

Самостоятельная работа студентов включает в себя решение задач, изучение лекционного материала, учебников, учебных пособий и иных материалов. Время и место самостоятельной работы выбираются студентами по своему усмотрению с учетом рекомендаций преподавателя.

Самостоятельную работу над дисциплиной следует начинать с изучения учебно-методического комплекса, который содержит основные требования к знаниям, умениям, навыкам. Необходимо также вспомнить рекомендации преподавателя, данные в ходе лекционных занятий или консультаций, затем приступить к изучению отдельных разделов и тем.

Подготовка к практическим занятиям:

- выполнить практические задания домашней работы;
- внимательно изучить материал лекций, относящихся к данному семинарскому занятию, ознакомиться с учебным материалом по учебнику и учебным пособиям;
- выписать основные термины;
- уяснить, какие учебные элементы остались для неясными и сформулировать вопросы, которые необходимо задать преподавателю на занятии или консультации;
- готовиться можно индивидуально, парами или в составе малой группы, последние являются эффективными формами работы.

Для подготовки используются презентации с материалами семинаров и консультации при работе в компьютерном классе.

3.1.2 Методическое обеспечение самостоятельной работы

Презентации с изучаемым материалом, консультации во время аудиторных занятий.

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

Презентация командных разработок, презентация своей роли и научно-практического вклада в командную разработку проекта, демонстрация знаний теоретического материала. Для получения промежуточной аттестации необходимо выполнить все практические задания и описать значимость собственных результатов в командном проекте.

Для получения итоговой оценки необходимо сдать зачет по курсу.

Итоговая оценка "зачтено" ставится при условии выполнения всех практических заданий в течении семестра, а также при активной работе в команде над проектом.

Итоговая оценка "не зачтено" ставится при отсутствии хотя бы двух практических заданий, а также при пассивной работе в команде над проектом. Преподаватель имеет право предоставить информацию о задолженностях студента в аттестационную комиссию.

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

Текущая успеваемость обучаемых контролируется степенью их активности на семинарах (выполнение докладов) и степенью их вовлеченности в процесс командной разработки (согласно выбранному заданию). Примерные темы докладов на семинарах и направления практических задач определены в п. 2.2.

3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса

Для оценки обучающимися содержания и качества учебного процесса используется анкета-отзыв установленная локальными актами СПбГУ.

3.2. Кадровое обеспечение

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

Знание предметной области, коммуникабельность

3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом

Нет специальных требований

3.3. Материально-техническое обеспечение

3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий

Наличие ПК, мультимедийного проектора, доски

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

Графический редактор, Microsoft Office 2013, Internet

3.3.3 Характеристики специализированного оборудования

Нет специальных требований

3.3.4 Характеристики специализированного программного обеспечения

Нет специальных требований

3.3.5 Перечень и объёмы требуемых расходных материалов

Нет специальных требований

3.4. Информационное обеспечение

3.4.1 Список обязательной литературы

1. К. Маннинг, П. Рагхаван, Х. Шютце. Введение в информационный поиск // М.: ИД «Вильямс». – 2011. – С. 528.

2. Ian H. Witten, Eibe Frank and Mark A. Hall Data Mining: Practical Machine Learning Tools and Techniques// 3rd Edition. — Morgan Kaufmann, 2011. — P. 629.

3.4.2 Список дополнительной литературы

1. G. Pearman, J. Goodwill. Pro .NET 2.0 Extreme Programming (Expert's Voice) // Apress. - 2006. - P. 344.
2. S. M. Weiss, N. Indurkha, T. Zhang, F. Damerau. Text Mining: Predictive Methods for Analyzing Unstructured Information // Springer. - 2010. – P. 237.
3. Классификация и кластер. Под ред. Дж. Вэн Райзина. М.: Мир. - 1980. - С. 390.
4. А. Коберн. Быстрая разработка программного обеспечения / Лори. - 2002. - С. 336.

3.4.3 Перечень иных информационных источников

<http://dl.acm.org/>, <http://books.google.com>, <http://www.elibrary.ru>

Раздел 4. Разработчики программы

Фамилия, имя, отчество	Учёная степень	Учёное звание	Должность	Контактная информация (служебный адрес электронной почты, служебный телефон)
Блеканов Иван Станиславович	к.т.н		доцент	i.blekanov@spbu.ru
Стученков А.Б.			старший преподаватель	a.stuchenkov@spbu.ru