

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Шевчик Андрей Павлович
Должность: Ректор
Дата подписания: 17.10.2023 13:53:54
Уникальный программный ключ:
476b4264da36714552dc83748d2961662babc012



МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Санкт-Петербургский государственный технологический институт
(технический университет)»
(СПбГТИ(ТУ))

УТВЕРЖДАЮ

Ректор _____ А.П. Шевчик

25 января 2022 г.

Рабочая программа дисциплины
ХЕМОИНФОРМАТИКА

Научная специальность
1.4.5 Хемоинформатика

Подготовка научных и научно-педагогических кадров в аспирантуре

Санкт-Петербург

2022

СОДЕРЖАНИЕ

1. Перечень планируемых результатов освоения дисциплины.....	4
2. Место дисциплины в структуре образовательной программы.....	4
3. Объем дисциплины	4
4. Содержание дисциплины.....	5
5. Порядок проведения промежуточной аттестации.....	8
6. Рекомендуемая литература	8
7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	8
8. Методические указания для аспирантов по освоению дисциплины.....	8
9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине.....	9
10. Материально-техническое обеспечение дисциплины.....	9
11. Особенности освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья	9

1. Перечень планируемых результатов освоения дисциплины

Цель изучения дисциплины – углубленное изучение наиболее важных и актуальных теоретических и практических вопросов, охватываемых паспортом специальности 1.4.5 Хемоинформатика, приобретение навыков использования научных методов и средств для решения теоретических и прикладных задач научной специальности, подготовка к сдаче кандидатского экзамена по специальности «Хемоинформатика».

Задачи изучения дисциплины:

- углубление и расширение теоретических знаний по хемоинформатике;
- овладение методами и средствами научного исследования в хемоинформатике;
- систематизация знаний в области хемоинформатики;
- подготовка к сдаче кандидатского экзамена по хемоинформатике.

В результате освоения образовательной программы аспирант должен продемонстрировать следующие результаты освоения дисциплины «Хемоинформатика»:

- способность применять углубленные знания в профессиональной деятельности в области хемоинформатики;
- способность адаптировать новое знание к узкопрофессиональной и междисциплинарной деятельности в области хемоинформатики;
- способность к самостоятельному построению и аргументированному представлению научной гипотезы;
- свободное владение всеми разделами хемоинформатики, умение ориентироваться в разнообразии методологических подходов.

2. Место дисциплины в структуре образовательной программы.

Дисциплина «Хемоинформатика» относится к образовательному компоненту программы аспирантуры и относится к обязательным элективным дисциплинам, направленным на подготовку к сдаче кандидатских экзаменов. Дисциплина изучается на 2 курсе в 3 и 4 семестрах.

Полученные в процессе изучения дисциплины «Хемоинформатика» знания, умения и навыки могут быть использованы в научно-исследовательской работе аспиранта.

3. Объем дисциплины.

Вид учебной работы	Всего, академических часов
	Очная форма обучения
Общая трудоемкость дисциплины (зачетных единиц/ академических часов)	5 / 180
Контактная работа с преподавателем:	40
Обзорно-установочные лекции и консультации	40
Самостоятельная работа	104
Форма промежуточной аттестации - кандидатский экзамен (4 сем.)	36

Рабочая программа дисциплины рассчитана на **5 ЗЕТ (180 час.)**, из них около 20% могут составлять аудиторные занятия, включая обзорно-установочные лекции, консультации с преподавателем. Основная часть работы аспиранта является самостоятельной и включает изучение рекомендованной преподавателем литературы, работу с источниками, подготовку к кандидатскому экзамену.

Обзорно-установочные лекции и консультации могут проводиться, в том числе, с использованием дистанционных образовательных технологий, электронного обучения.

4. Содержание дисциплины.

4.1. Разделы дисциплины и виды занятий.

№ п/п	Наименование раздела дисциплины	Обзорно-установочные лекции, консультации акад. часы	Самостоятельная работа, акад. Часы
1	Представления молекул	8	20
2	Химические базы данных	8	20
3	Дескрипторы	8	20
4	Хеометрика	8	20
5	Построение моделей «структура-свойство»	8	24

4.2. Обзорно-установочные лекции

№ разд. дисц.	Наименование тем обзорно-установочных лекций	Объем, акад. часы
1	Химическая номенклатура как линейное представление. Линейные представления Висвессера (WLN). Линейные представления SMILES. Линейное представление реакций SMIRKS. Представление шаблонов для спецификации молекулярных фрагментов SMART. Линейные представления SLN. Идентификатор InChI. Векторное представление молекулы. Битовое представление молекулы. Матричное и табличное представления. Форматы представлений.	8
2	Классификация баз данных. Структурный поиск в химических базах данных. Алгоритмы хэширования. Виды поиска. Поиск в базах данных химических реакций. Поиск в базах данных трёхмерных структур. Изоморфизм, автоморфизм и симметрия графов. Алгоритмы прохождения деревьев решений. Виды поиска на графах. Важнейшие химические базы данных: CAS/SCIFINDER, Reaxys, ChemSpider, PubChem, ChEMBL. ZINC, BINDINGDB.	8

№ разд. дисц.	Наименование тем обзорно-установочных лекций	Объем, акад. часы
3	Классификация дескрипторов. Топологические, трёхмерные (3D), геометрические, координатные, топографические, фрагментарные виды дескрипторов. Дескрипторы на основе: цепочек атомов, строк WLN и SMILES, центрированных, библиотечных и случайных фрагментов, максимальных общих подструктур. Фармакофорные дескрипторы. Константы заместителей. Физико-химические дескрипторы. Квантово-химические дескрипторы. Дескрипторы молекулярных орбиталей. Дескрипторы электронного распределения. Энергетические и термодинамические дескрипторы. Дескрипторы молекулярных полей.	8
4	Многомерность данных. Скрытые структуры данных. Математическая предобработка данных. Линейный метод главных компонент (МГК). Интерпретация графиков нагрузок и счетов. Алгоритмы вычисления главных компонент. Особенности построения моделей МГК. Множественная линейная регрессия. Метод проекций на латентные структуры (ПЛС). ПЛС-модели. Нелинейные МГК. Виды ядер-функций. Ядро-МГК. Метод <i>k</i> ближайших соседей. Искусственные нейронные сети. Метод опорных векторов. Деревья принятия решений.	8
5	Нормализация специфических хемотипов, резонансных форм и таутомеров. Предобработка химических структур. Удаление смесей, неорганических и металлоорганических соединений. Конвертация структур, удаление солей и выбор состояния ионизации. Выявление дубликатов. Оверфиттинг и принцип оптимальной сложности модели. Принципы отбора дескрипторов. Сравнение качества моделей. Область применимости моделей. Принцип Сетубала (ОЭСР). Рекомендации Унегра-Ганча. Рекомендации Тропши по практике построения, валидации и применения моделей QSAR.	8

4.3. Самостоятельная работа аспирантов.

№ разд. дисц.	Наименование тем для самостоятельной работы	Объем, акад. часы

№ разд. дисц.	Наименование тем для самостоятельной работы	Объем, акад. часы
1	Химическая номенклатура как линейное представление. Линейные представления Висвессера (WLN). Линейные представления SMILES. Линейное представление реакций SMIRKS. Представление шаблонов для спецификации молекулярных фрагментов SMART. Линейные представления SLN. Идентификатор InChI. Векторное представление молекулы. Битовое представление молекулы. Матричное и табличное представление. Форматы представлений.	20
2	Классификация баз данных. Структурный поиск в химических базах данных. Алгоритмы хэширования. Виды поиска. Поиск в базах данных химических реакций. Поиск в базах данных трёхмерных структур. Изоморфизм, автоморфизм и симметрия графов. Алгоритмы прохождения деревьев решений. Виды поиска на графах. Важнейшие химические базы данных: CAS/SCIFINDER, Reaxys, ChemSpider, PubChem, ChEMBL. ZINC, BINDINGDB.	20
3	Классификация дескрипторов. Топологические, трёхмерные (3D), геометрические, координатные, топографические, фрагментарные виды дескрипторов. Дескрипторы на основе: цепочек атомов, строк WLN и SMILES, центрированных, библиотечных и случайных фрагментов, максимальных общих подструктур. Фармакофорные дескрипторы. Константы заместителей. Физико-химические дескрипторы. Квантово-химические дескрипторы. Дескрипторы молекулярных орбиталей. Дескрипторы электронного распределения. Энергетические и термодинамические дескрипторы. Дескрипторы молекулярных полей.	20
4	Многомерность данных. Скрытые структуры данных. Математическая предобработка данных. Линейный метод главных компонент (МГК). Интерпретация графиков нагрузок и счетов. Алгоритмы вычисления главных компонент. Особенности построения моделей МГК. Множественная линейная регрессия. Метод проекций на латентные структуры (ПЛС). ПЛС-модели. Нелинейные МГК. Виды kern-функций. Kern-МГК. Метод <i>k</i> ближайших соседей. Искусственные нейронные сети. Метод опорных векторов. Деревья принятия решений.	20
5	Нормализация специфических хемотипов, резонансных форм и таутомеров. Предобработка химических структур. Удаление смесей, неорганических и металлоорганических соединений. Конвертация структур, удаление солей и выбор состояния ионизации. Выявление дубликатов. Оверфитинг и принцип оптимальной сложности модели. Принципы отбора дескрипторов. Сравнение качества моделей. Область применимости моделей. Принцип Сетубала (ОЭСР). Рекомендации Унегра-Ганча. Рекомендации Тропши по практике построения, валидации и применения моделей QSAR.	24

5. Порядок проведения промежуточной аттестации

Промежуточная аттестация по дисциплине проводится в форме кандидатского экзамена в соответствии с выбранной специальностью.

Экзамен предусматривает выборочную проверку освоения предусмотренных результатов обучения по дисциплине и комплектуется вопросами, представленными в программе кандидатского экзамена по научной специальности 1.4.5 Хемоинформатика.

6. Рекомендуемая литература

а) печатные издания

1. Петров, А.А. Органическая химия / А.А. Петров, Х.В. Бальян., А.Т. Трощенко. – 5-е изд. перераб. и доп. – Санкт-Петербург : Иван Федоров, 2015. – 624 с. - ISBN 978-5-903034-99-4.
2. Денисов В.Я. Органическая химия. Учебник /Денисов В.Я., Мурышкин Д.Л, Чуйкова Т.В. – Москва : – Высшая Школа, 2009. – 544 с. – ISBN 9785-06-005743-0.
3. Ключинский, С.А. Информационные ресурсы по органической химии в Интернете и графические инструменты (редакторы химических структур) для работы с ними / учебное пособие / С.А. Ключинский. – Санкт-Петербург : СПбГТИ(ТУ), 2013. – 68 с.

б) электронные издания

1. Якубик, Д. Г. Химическая информатика : учебное пособие / Д. Г. Якубик. – Кемерово : КемГУ, 2021. – 79 с. – ISBN 978-5-8353-2734-8. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/173539> (дата обращения: 26.01.2022). – Режим доступа: по подписке.

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.

1. Библиотека Санкт-Петербургского государственного технологического института (технического университета) университета - <http://bibl.lti-gti.ru>
2. Российская государственная библиотека - www.rsl.ru
3. Российская национальная библиотека - www.nlr.ru
4. Библиотека Академии наук - www.rasl.ru
5. Библиотека по естественным наукам РАН - www.benran.ru
6. Всероссийский институт научной и технической информации (ВИНИТИ) - www.viniti.ru
7. Государственная публичная научно-техническая библиотека - www.gpntb.ru
8. Научная электронная библиотека eLIBRARY.RU - elibrary.ru
9. Реферативная база данных научных публикаций Web of Science - webofknowledge.com
10. Электронно-библиотечная система "Лань" <http://e.lanbook.com>

8. Методические указания для аспирантов по освоению дисциплины.

Методические указания для аспирантов по организации самостоятельной работы по дисциплине, включая перечень тем самостоятельной работы, размещены в электронной информационно-образовательной среде СПбГТИ(ТУ) на сайте Медиа: <http://media.technolog.edu.ru>

Планирование времени, необходимого на изучение данной дисциплины, лучше всего осуществлять на оба семестра, предусматривая при этом регулярное повторение пройденного материала.

Основными условиями правильной организации учебного процесса для аспирантов являются:

- плановость в организации учебной работы;
- серьезное отношение к изучению материала;
- постоянный самоконтроль.

В ходе обзорно-установочных лекций преподаватель излагает и разъясняет основные, наиболее сложные понятия темы, а также связанные с ней теоретические и практические проблемы, дает рекомендации на выполнение самостоятельной работы.

В ходе лекций аспирантам рекомендуется:

- вести конспектирование учебного материала;
- обращать внимание на категории, формулировки, раскрывающие содержание тех или иных определений или понятий, научные выводы и практические рекомендации по их применению;
- задавать преподавателю уточняющие вопросы.

Самостоятельная работа – ключевой аспект освоения аспирантом дисциплины «Хемоинформатика», основывающийся на понимании материала, излагаемого в ходе обзорно-установочных лекций, самостоятельном поиске, подборе и обработке информации. При этом значительную часть необходимых для освоения курса данных необходимо будет найти в научной литературе.

9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине.

9.1. Информационные технологии.

В учебном процессе по данной дисциплине предусмотрено использование информационных технологий:

- чтение лекций с использованием слайд-презентаций;
- взаимодействие с аспирантом посредством электронно-информационной образовательной среды.

9.2. Программное обеспечение.

Microsoft Windows 10 / 11; Microsoft Office (Excel, Word, PowerPoint).

10. Материально-техническое обеспечение дисциплины.

Для ведения лекций используется аудитория, оборудованная средствами оргтехники, Материально-техническое обеспечение дисциплины: доступ к фондам учебных пособий, библиотечным фондам с периодическими изданиями по соответствующим темам, наличие компьютеров, подключенных к сети Интернет и оснащенных средствами медиапрезентаций (медиакоммуникаций).

11. Особенности освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья.

Для инвалидов и лиц с ограниченными возможностями учебные процесс осуществляется в соответствии с Положением об организации учебного процесса для обучения инвалидов и лиц с ограниченными возможностями здоровья СПбГТИ(ТУ), утвержденным ректором 28.08.2014г.