



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

УТВЕРЖДАЮ

Ректор

Шев

А.Н. Шевчик

« 25 » *сентября* 2022 г.



ПРОГРАММА
вступительных испытаний для приема на обучение по программе
подготовки научных и научно-педагогических кадров в аспирантуре

По дисциплине

ХЕМОИНФОРМАТИКА

Научная специальность

1.4.5 – Хемоинформатика

Санкт-Петербург

2022

СОДЕРЖАНИЕ

| | | |
|---|---|---|
| 1 | Рекомендуемая структура экзамена | 3 |
| 2 | Разделы дисциплины, рассматриваемые в ходе экзамена..... | 3 |
| 3 | Вопросы к вступительному экзамену..... | 5 |
| 4 | Учебные издания..... | 6 |
| 5 | Методические указания по подготовке к вступительному экзамену..... | 7 |

1. Рекомендуемая структура экзамена

1.1. Письменные ответы на три вопроса из списка экзаменационных вопросов.

1.2. Беседа с членами экзаменационной комиссии по этим вопросам и вопросам, связанным со специальностью и будущим научным исследованием.

2. Разделы дисциплины, рассматриваемые в ходе экзамена

2.1. Представления молекул

2.1.1. Линейные представления

Химическая номенклатура как линейное представление. Линейные представления Висвессера (WLN). Линейные представления SMILES. Правила SMILES. Канонические представления SMILES. Указание стереохимии SMILES. Линейное представление реакций SMIRKS. Представление шаблонов для спецификации молекулярных фрагментов SMART. Линейные представления SLN. Идентификатор InChI.

2.1.2. Представления молекул в виде графов

Векторное представление молекулы. Битовое представление молекулы. Матричное представление. Матрица смежности. Матрица расстояний. Матрица инцидентности. Матрица связей. Матрица связей-электронов. Табличное представление.

2.1.3. Трёхмерные представления молекулы

Форматы MDL. Структурный формат MOL-формат. Формат данных SDF-формат. Реакционный формат RXN-формат. Формат реакций и данных RDF-формат. Молекулярный формат Sybyl mol2. Формат базы данных белков PDB. Формат данных CML.

2.2. Химические базы данных

2.2.1. Структурный поиск в химических базах данных

Общие сведения о базах данных. Классификация баз данных. Структурный поиск в химических базах данных. Виды поиска, основанного на структуре. Поиск по структуре. Алгоритм хэширования Гаштайгера-Иленфельда. Поиск по подструктуре. Использование скринов. Изоморфное вложение графов. Алгоритм Ульмана. Поиск по суперструктуре. Поиск по сходству. Поиск по структурам Маркуша. Поиск в базах данных химических реакций. Поиск в базах данных трёхмерных структур. Жёсткий и гибкий трёхмерный поиск.

2.2.2. Алгоритмы на графах.

Основные определения теории графов. Изоморфизм, автоморфизм и симметрия графов. Алгоритмы прохождения деревьев решений. Поиск

компонент связности. Поиск цепочек и циклов в графе. Поиск клик в графах и максимальных общих подграфов. Алгоритм поиска кратчайших путей в графах. Алгоритм Флойда-Воршала для нахождения матрицы расстояний в графе.

2.2.3. Важнейшие химические базы данных

CAS/SCIFINDER. Reaxys. ChemSpider. PubChem. ChEMBL. ZINC. BINDINGDB. Кембриджская структурная база данных. Protein Data Bank. Sc-PDB. Базы данных профайлинга химических соединений.

2.3. Дескрипторы

Классификация дескрипторов. Топологические дескрипторы. Топологические индексы, построенные на матрице смежности. Топологические индексы на основе подсчёта путей в графе. Дескрипторы на основе матрицы расстояний. Дескрипторы на основе рёберной матрицы смежности. Двумерные автокорреляции. Индексы информационной теории. Трёхмерные (3D) дескрипторы. Геометрические дескрипторы. Координатные дескрипторы. Топографические дескрипторы. Дескрипторы поверхности. Дескрипторы на основе площади поверхности молекулы. Дескрипторы, основанные на описании распределения характеристике по поверхности молекулы. Дескрипторы объёма и формы. Фрагментарные дескрипторы. Дескрипторы на основе цепочек атомов. Дескрипторы на основе строк WLN и SMILES. Дескрипторы на основе атом-центрированных фрагментов. Дескрипторы на основе связь-центрированных фрагментов. Дескрипторы на основе топологических мультиплетов. Дескрипторы на основе максимальных общих подструктур. Дескрипторы на основе базовых подграфов. Дескрипторы на основе подграфов, найденных при помощи интеллектуального анализа графов. Дескрипторы на основе случайных фрагментов. Дескрипторы на основе библиотечных фрагментов. Фармакофорные дескрипторы. Дескрипторы на основе топологических фармакофоров. Дескрипторы на основе пространственных фармакофоров. Константы заместителей. Стерические константы заместителей. Электронные константы заместителей. Физико-химические дескрипторы. Молекулярная масса. Дескрипторы гидрофобности (липофильности). Молекулярная рефракция. Дескрипторы водородной связи. Дескрипторы растворителей. Квантово-химические дескрипторы. Дескрипторы молекулярных орбиталей. Дескрипторы электронного распределения. Энергетические и термодинамические дескрипторы. Дескрипторы молекулярных полей. Дескрипторы молекулярного подобия.

2.4. Построение моделей «структура-свойство»

2.4.1. Предобработка данных

Предобработка химических структур. Удаление смесей, неорганических и металлоорганических соединений. Конвертация структур, удаление солей и выбор состояния ионизации. Нормализация

специфических хемотипов, резонансных форм и таутомеров. Выявление дубликатов. Ручная проверка. Математическая предобработка данных.

2.4.2. Общие принципы построения моделей

Метод наименьших квадратов и оверфиттинг. Оверфиттинг и принцип оптимальной сложности модели. Принципы отбора дескрипторов. Проблемы, связанные с отбором дескрипторов.

2.4.3. Методы машинного обучения

Множественная линейная регрессия. Метод частичных наименьших квадратов. Метод k ближайших соседей. Искусственные нейронные сети. Метод опорных векторов. Деревья принятия решений.

2.4.4. Валидация моделей

Сравнение качества моделей. Область применимости моделей. Принцип Сетубала (ОЭСР). Рекомендации Унегра-Ганча. Рекомендации Тропси по практике построения, валидации и применения моделей QSAR.

3. Вопросы к вступительному экзамену

3.1. Вопросы по разделу «Представления молекул»

- 1) В чем заключаются особенности представления молекул в хемоинформатике.
- 2) Требования к представлениям молекул, пригодным для использования в автоматизированных системах.
- 3) Классификации молекулярных представлений.
- 4) Основные классы битовых представлений.
- 5) Преимущества и недостатки молекулярных отпечатков.

3.2. Вопросы по разделу «Химические базы данных»

- 6) Классификация химических баз данных.
- 7) Виды структурного поиска.
- 8) Основные фактографические химические базы данных.
- 9) Основные структурные химические базы данных.

3.3. Вопросы по разделу «Дескрипторы»

- 10) Требования, которым удовлетворять молекулярные дескрипторы.
- 11) Признаки классификации дескрипторов.

3.3. Вопросы по разделу «Построение моделей «структура-свойство»

- 12) Химическая предобработка данных: отбор данных и стандартизация.
- 13) Основные способы машинного обучения.
- 14) Валидация данных.

15) Математическая предобработка данных: стандартизация, шкалирование, нормализация. Случайная корреляция и борьба с ней.

4. Учебные издания

а) печатные учебные издания:

1. Петров, А.А. Органическая химия / А.А. Петров, Х.В. Бальян., А.Т. Трощенко. – 5-е изд. перераб. и доп. – СПб.: Иван Федоров, 2015. – 624 с.

2. Денисов В.Я. Органическая химия. Учебник /Денисов В.Я., Мурышкин Д.Л, Чуйкова Т.В. – М.: – Высш. Школа, 2009. – 544 с.

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. – СПб.: СПбГТИ(ТУ), 2013. – 44с.

б) электронные учебные издания:

4. 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.

в) вспомогательные печатные и электронные источники:

5. Маджидов, Т. И. Введение в хемоинформатику. Компьютерное представление химических структур [Текст] : учебное пособие / Т. И. Маджидов [и др. ; науч. ред. – Г. А. Чмутова]. – Казань : Казанский ун-т, 2013. – 174 с. – ISBN 978-5-00019-131-6.

6. Баскин, И. И. Введение в хемоинформатику. Химические базы данных [Текст] : учебное пособие / И. И. Баскин, Т. И. Маджидов, А. А. Варнек ; Казанский федеральный ун-т. – Казань : Изд-во Казанского ун-та, 2015. – 188 с. – ISBN 978-5-00019-429-34.

7. Баскин, И. И. Введение в хемоинформатику. Моделирование "структура-свойство" [Текст] : учебное пособие / И. И. Баскин, Т. И. Маджидов, А. А. Варнек ; Казанский федеральный ун-т. – Казань : Изд-во Казанского ун-та, 2015. – 304 с. – ISBN 978-5-00019-442-3.

8. Баскин, И. И. Введение в хемоинформатику. Методы машинного обучения [Текст] : учебное пособие / И. И. Баскин, Т. И. Маджидов, А. А. Варнек ; Казанский федеральный ун-т. – Казань : Изд-во Казанского ун-та, 2016. – 330 с. – ISBN 978-5-00019-695-3.

9. Баскин, И. И. Введение в хемоинформатику. Информатика химических реакций [Текст] : учебное пособие / И. И. Баскин, Т. И. Маджидов, А. А. Варнек ; Казанский федеральный ун-т. – Казань : Изд-во Казанского ун-та, 2017. – 244 с. – ISBN 978-5-00019-907-7.

10. Баскин, И. И. Введение в хемоинформатику. Химическое пространство и дизайн библиотек [Текст] : учебное пособие / И. И. Баскин, Т.

И. Маджидов, А. А. Варнек ; Казанский федеральный ун-т. – Казань : Изд-во Казанского ун-та, 2019. – 240 с. – ISBN 978-5-00130-174-5.

11. Gasteiger, J. Chemoinformatics: A Textbook / J. Gasteiger, T. Engel. – Weinheim: Wiley-VCH, 2003. – 649 p. – ISBN 978-3527306817.

12. Faulon, J. L. Handbook of Chemoinformatics Algorithms / J. L. Faulon, F. Bender. – Boca Raton: CRC Press, 2010. – 454 p. – ISBN 978-1420082920.

13. Todeshini, R. Molecular Descriptors for Chemoinformatics. Vol. I: Alphabetical listing / R. Todeshini, V. Consonni. – Weinheim: Wiley-VCH, 2009. – 967 p. – ISBN 978-3527318520.

5. Методические указания по подготовке к вступительному экзамену

При подготовке к вступительному экзамену поступающим в аспирантуру необходимо ориентироваться на разделы, посвященные методам и программным средствам математического моделирования и оптимизации технических объектов, методам планирования и обработки результатов экспериментов, методам численного решения уравнений математических моделей различных классов, методам и технологиям разработки программных комплексов для моделирования и исследования технических объектов, которые были изучены в ходе освоения соответствующих учебных дисциплин в бакалавриате и магистратуре или специалитете. В ходе подготовки к вступительному экзамену следует выполнить анализ современной литературы и электронных ресурсов (в том числе изданий электронно-библиотечных систем, ресурсов информационно-телекоммуникационной сети «Интернет»), приведенных в настоящей программе, и общеобразовательную проработку изложенных в ней разделов дисциплины. Необходимо прочитать рекомендованные издания и при необходимости составить краткий конспект основных положений, терминов, сведений, требующих запоминания и являющихся основополагающими в каждом разделе дисциплины.

Для расширения и углубления знаний по дисциплине рекомендуется использовать следующие Интернет-ресурсы:

Фундаментальная библиотека Санкт-Петербургского государственного технологического института (технического университета) университета (URL: <http://bibl.lti-gti.ru>);

Российская государственная библиотека (URL: <https://www.rsl.ru>);

Российская национальная библиотека (URL: <http://nlr.ru>);

Библиотека Академии Наук (URL: <http://www.ras.ru>);

Государственная публичная научно-техническая библиотека России (URL: <https://www.gpntb.ru>);

Всероссийский институт научной и технической информации Российской академии наук (ВИНИТИ РАН) (URL: <http://www.viniti.ru>);

Научная электронная библиотека eLIBRARY.RU (URL: <https://elibrary.ru>);

информационно-поисковая система Интернет-портала Федерального института промышленной собственности (URL: <https://new.fips.ru/iiss>);

образовательный математический сайт (URL: <https://exponenta.ru>);

федеральный портал «Российское образование» (URL: <https://edu.ru>);

Российский портал открытого образования (URL: <https://openedu.ru>).