ПРОГРАММА-МИНИМУМ
кандидатского экзамена по специальности
1.2.3 «Теоретическая информатика, кибернетика»
по физико-математическим и техническим наукам

Программа разработана в МГУ имени М.В. Ломоносова с привлечением специалистов механико-математического факультета, факультета вычислительной математики и кибернетики, института проблем информационной безопасности, института механики.

Целью экзамена кандидатского минимума является оценка уровня профессиональных знаний соискателя ученой степени кандидата наук и уровня его подготовки к выполнению самостоятельной исследовательской работы в рамках научных направлений выбранной специальности. Настоящая программа соответствует актуальным профессиональным компетенциям в области теоретической информатики. Программа основана на материалах учебных курсов по теоретической информатике, которые представлены для студентов и аспирантов в рамках учебных программ факультетов МГУ имени М.В. Ломоносова. Программа предполагает владение необходимыми элементами математического аппарата таких дисциплин, как дискретная математика, теория чисел, математическая логика и теория алгоритмов, теория вероятностей и математическая статистика.

Общие положения

Информатика как наука, изучающая информацию и ее свойства в естественных, искусственных и гибридных системах. Место информатики в системе наук. Информатика как обрабатывающая информацию отрасль индустрии и инфраструктурная область, ее роль и значение в ускорении научно-технического прогресса.

Понятие информационного продукта и информационной услуги. Классификация информационных продуктов и услуг. Жизненный цикл информационного продукта. Экономика информационных сетей. Методы управления производством и распределением информационных продуктов.

Информационные ресурсы. Принципы оценки информации как ресурса общества и объекта интеллектуальной собственности. Проблемы правового регулирования научной интеллектуальной собственности. Государственная политика в области защиты информационных ресурсов общества.

Концептуальные модели

Машинное обучение. Задача машинного обучения. Объекты и признаки. Основные понятия: метод обучения, функционал качества, обобщающая способность, скользящий контроль. Алгоритмы классификации: C4.5, анализ формальных понятий, метод опорных векторов, k ближайших соседей, Байесовские классификаторы, AdaBoost, скрытые модели Маркова, метод условных случайных полей. Факторный анализ. Алгоритмы кластеризации: k-средних, самоорганизующиеся карты Кохонена, графовые алгоритмы, алгоритмы, основанные на плотности (DBSCAN), методы нечеткой кластеризации, иерархическая кластеризация.

Методы анализа текстовых данных. Приложения задач анализа текстовых данных: поиск схожих объектов, кластеризация, извлечение данных, выявление трендов. Алгоритмы выделения именованных сущностей, гиперонимов, описаний объектов. Разрешение неоднозначности. Алгоритмы выявления ассоциативных правил: A-Priori, PCY, SON, алгоритм Тойвонена, FP-growth.

Приближенные алгоритмы дискретной оптимизации. Примеры оптимизационных задач: минимальное трансверсальное множество, минимальное покрывающее множество. Жадные алгоритмы. Сведение к задаче линейного программирования. Алгоритмы со случайным выбором.

Алгоритмы анализа социальных сетей. Формализация понятия сообщества в социальной сети. Алгоритмы нахождения сообществ (Гирвана-Ньюмана и др.), SimRank. Оценка авторитетности ресурсов. Алгоритм PageRank и его вариации.

Методы представления знаний. Базы знаний. Общие принципы моделирования окружающей среды и мышления человека. Методы представления знаний: классификационные, тезаурусные, основанные на отношениях, семантические сети и фреймы, продукционные и непродукционные.

Онтологии. Введение в дескриптивную логику. Семейства логик. Характерные задачи. Логический вывод. Онтологии. Языки описания онтологий. Языки запросов к онтологиям.

Семантическая паутина. Введение в Семантическую паутину (Semantic Web). Краткая история развития Всемирной паутины. Основные технологии Semantic Web. Структура SW-приложения. Два подхода к реализации. Проект Linking Open Data. Перспективы развития Semantic Web.

Распознавание образов

Задача распознавания образов. Основные подходы: геометрический, вероятностный и комбинаторно-логический. Примеры задач распознавания. Геометрический подход. Линейные процедуры распознавания. Перцептроны. Теорема Новикова. Метод потенциальных функций.

Вероятностный подход. Процедура Байеса. Метод обобщенного портрета. Условия кластеризации. Основные процедуры построения кластеров. Метод скрытых марковских процессов.

Комбинаторно-логический подход. Линейные процедуры и информационные веса. Условия эффективности распознавания. Тесовые процедуры распознавания. Алгоритм голосования. Оценки длины минимальных тестов.

Оценки числа тупиковых тестов для почти всех таблиц. Асимптотически оптимальный алгоритм построения множества коротких тестов. Полиномиальный характер решающих правил распознавания.

Управление данными и информационный поиск

Модели данных. Понятие модели данных. Иерархическая, сетевая модели данных, сравнительный анализ, противоречия и парадоксы. Реляционная модель данных. Экземпляры отношений, домены, атрибуты. Операции над отношениями: селекция, проекция, естественное соединение. Понятие реляционной полноты языка манипулирования данными. Модель данных «сущность-связь». Модели данных в No-SQL и New-SQL системах.

Базы данных. Независимость программ и данных. Интегрированное использование данных. Непротиворечивость данных. Целостность и защита данных. Структуры БД. Администрирование баз данных. Типы пользователей. Администратор БД. Понятие концептуальной, логической, физической структуры БД. Представления пользователей и подсхемы. Понятие о словарях данных, языках описания и манипулирования данными. БД и файловые системы.

Системы управления базами данных. Состав и структура. Типовые функции СУБД: хранение, поиск и модификация данных; обеспечение параллельного доступа к данным; преобразование данных; словарное обеспечение БД; обеспечение целостности и непротиворечивости данных. Язык описания и манипулирования данными SQL. Транзакции,

целостность данных. Типовая структура СУБД: ядро, интерпретатор/компилятор запросов, менеджеры ресурсов, буферов, протоколирования и восстановления, транзакций, планировщик заданий. No-SQL и New-SQL СУБД. Теорема CAP.

Полнотекстовые БД. Физическая и логическая структура. Файл полного текста. Частотный словарь, инверсный файл. Положительный и отрицательный словари. Стандартные строки и словосочетания, включаемые в частотный словарь. Описание БД. Обработка текстов при загрузке БД. Понятие экспорта импорта документов-данных.

Построение распределенных баз данных. Основные способы построения распределенных приложений. Распределенные транзакции. Механизмы репликации.

Обеспечение безопасности в СУБД. Методы шифрования протоколов обмена данными. Построение виртуальных баз данных. Защита структур данных. Распределение прав доступа к данным. Политики распределения ресурсов в СУБД.

Обработка слабоструктурированных данных. Графовые и древовидные модели данных. Понятие путевого запроса. Языки запросов XPath и XQuery. Методы выполнения запросов.

Критерии оценки информационных технологий и систем. Оценки качества поиска (полнота, точность и др.). Скалярные и векторные оценки. Смешанные критерии (полезная работа, корреляционный критерий, свертки и пр.). Рабочие характеристики информационно-поисковых систем (ИПС) в различных координатах. Вероятностная модель ИПС. Теоретико-множественная модель ИПС. Оптимизация режима ИПС. Линейное представление документов, запросов, индексирования, поиска.

Методы выполнения и оптимизации локальных и распределенных запросов. План выполнения запроса. Методы доступа к данным. Синтаксическая и стоимостная оптимизация. Адаптивные методы вычисления запросов.

Виртуальная интеграция данных. Задачи интеграции данных. Синтаксическая и семантическая неоднородность. Методы сопоставления схем данных. Перезапись запросов.

Элементы теории информации

Неопределенность и информация. Кодирование информации. Алфавитное кодирование. Теорема Маркова. Понятие энтропии стохастического источника. Измерение количества информации по Шеннону. Аксиоматика Шеннона-Хартли. Корректирующие свойства кодов. Коды Хэмминга, коды БЧХ. Теорема Шеннона о передаче при наличии помех.

Алгоритмический подход к понятию количества информации. Сложность конечного объекта по А.Н. Колмогорову. Существование оптимального способа описания. Количество информации. Сложность по Колмогорову и ее связь с шенноновской энтропией.

Элементы математической логики и теории алгоритмов

Логика первого порядка. Интерпретация формул. Нормальная форма Сколема. Проблема выполнимости. Теорема Черча. Метод резолюций.

Модели теории алгоритмов. Вычислимость. Интуитивное и формализованное понятие алгоритма. Машины Тьюринга, нормальные алгорифмы Маркова, частично-рекурсивные функции. Тезис Черча. Алгоритмически неразрешимые проблемы. Универсальный алгоритм.

Вычислимые и гёделевы (главные) нумерации вычислимых функций. Теорема Райса о неразрешимости проблемы распознавания любого нетривиального свойства вычислимой функции по ее описанию.

Сложность алгоритмов. Классы сложности P и NP, полиномиальная сводимость задач. Теория NP-полноты. Основные NP-полные задачи. Метод ветвей и границ и динамическое

программирование в комбинаторных алгоритмах. Градиентные и другие приближенные алгоритмы для NP-полных задач. Теоремы о соотношении сложности и точности приближенных алгоритмов для NP-полных задач. Метод рекурсии и оценки сложности алгоритмов, использующих рекурсию. Алгоритм быстрого преобразования Фурье. Принцип жадности. Теорема Радо-Эдмонса об условиях оптимальности жадного алгоритма.

Формальные языки и конечные автоматы

Детерминированные и недетерминированные конечные автоматы. Регулярные языки и регулярные выражения. Свойства регулярных языков. Лемма о накачке для регулярных языков и ее применение. Теорема Клини. Эквивалентные состояния и минимизация автоматов. Структурные автоматы. Алгоритмическая неразрешимость проблемы полноты для автоматов.

Понятие формального языка. Примеры формальных языков. Задание языков конечными автоматами. Порождение языков формальными грамматиками. Операции над языками.

Понятие вывода в формальной грамматике. Язык, порождаемый грамматикой. Линейные и автоматные грамматики и их свойства. Эквивалентность автоматных грамматик и конечных автоматов.

Контекстно-свободные грамматики (КС-грамматики) и контекстно-свободные языки (КС-языки). Деревья разбора. Автоматы с магазинной памятью (МП-автоматы). Эквивалентность МП-автоматов и КС-грамматик. Свойства КС-языков.

Нормальная форма Бэкуса-Наура. Примеры для арифметических выражений. Понятие о LL и LR разборах. Процедура LR(1) разбора выражений, соответствующих формальной грамматике. Операции сдвига и свертки.

Комбинаторные методы информатики

Принципы комбинаторики. Классификационные задачи. Учет различимости предметов и классов. Числа Белла и Стирлинга-1. Методы перечисления.

Выпуклый анализ и математическое программирование

Основные понятия выпуклого анализа и формулы выпуклого исчисления. Торема Фань Цзы. Теоремы об отделимости. Теорема Вейля. Теоремы о субдифференциале и об очистке. Принцип Лагранжа для выпуклых задач. Теорема Куна-Таккера. Операции Минковского над выпуклыми множествами.

Теоремы двойственности в выпуклом программировании. Теоремы двойственности и симплекс метод в линейном программировании. Транспортная задача и задача о назначении. Основные алгоритмы

Задача об оптимальном перемещении масс. Двойственность Монжа-Канторовича.

Алгоритмы поиска решений гладких, выпуклых и вариационных экстремальных задач. Метод центрированных сечений, метод эллипсоидов, метод симплексов. Метод внутренней точки (самосогласованные барьеры). Градиентный метод и его обобщения. Полуопределенное программирование.

Программные и технические основы

Классификация языков программирования. Структурное программирование. Объектно-ориентированный подход. Императивное, функциональное и логическое программирование. Статическая и динамическая типизация. Компиляция и интерпретация программ.

Язык программирования Си. Организация программы. Аргументы командной строки. Схема трансляции программ на языке Си. Препроцессор и компоновщик. Типы данных. Переменные. Константы. Массивы. Размещение массивов в памяти. Строки. Арифметические выражения. Логические выражения. Операторы if, while, for, do while, switch. Функции. Передача параметров и возврат значений. Область видимости и время существования переменных. Глобальные и локальные переменные. Рекурсия. Ввод-вывод. Работа с файлами.

Представление чисел. Представление целых чисел. Представление вещественных чисел. Стандарт IEEE 754. Специальные значения. Порядок байт (big endian, little endian).

Язык программирования Haskell. Строгая статическая типизация. Автоматический вывод типов. Неизменяемость данных. Ленивые вычисления. Функции (чистые функции, λ-функции, функции высших порядков). Рекурсия. Базовые типы данных. Алгебраические типы данных. Типы классов. Полиморфизм. Списки и операции над ними. Оператор if-else. Сопоставление с образцом. Ввод/вывод, монада IO.

Операционные системы. Архитектура ЭВМ. Ядро операционной системы. Файловая система. Управление памятью. Процессы, группы процессов. Планирование процессов. Межпроцессное взаимодействие (сигналы, каналы, разделяемая память, сокеты). Потоки, синхронизация потоков.

Архитектура TCP/IP. Модель OSI. Маршрутизация и топология сети. Маршрутизация в IP. Протокол UDP. Протокол TCP (установление и завершение соединения, трехэтапное рукопожатие, диаграмма состояний). Элементарные сокеты TCP (пример приложения клиент/сервер). Архитектура WWW. Протокол HTTP.

Принципы построения распределенных приложений. Основные понятия: открытость, совместное использование ресурсов, конкуренция, масштабируемость, отказоустойчивость, прозрачность. Базовые распределенные алгоритмы: выбор ведущего процесса, взаимное исключение, достижение согласия, многоадресная передача.

Математические методы анализа программных систем. Формальные модели описания систем: конечные автоматы, сети Петри, линейные временные логики. Алгоритмические задачи проверки выполнимости свойств модели.

Принципы создания информационных систем в сети Интернет. Клиент-серверная и многоуровневая архитектура программных систем. Особенности Web-приложений. Шаблоны проектирования сервера приложений.

Информационная безопасность

Комплексный подход к обеспечению информационной безопасности (ИБ). Выделение уровней комплексного подхода к обеспечению ИБ: законодательный, административно-организационный, операционный и программно-технический. Обзор мер и методов на уровнях комплексного подхода к обеспечению ИБ.

Модели угроз и нарушителей. Основные принципы построения систем защиты информации в компьютерных системах. Модель угроз и модель нарушителя. Распространенные угрозы и атаки. Проектирование и выбор архитектуры системы защиты. Автоматизированные системы в защищенном исполнении.

Аутентификация. Цифровая подпись. Идентификация и аутентификация в компьютерных системах. Общие положения создания систем и механизмов идентификации и аутентификации. Примеры реализации методов идентификации и аутентификации в операционных системах, в веб-приложениях.

Криптография и криптоанализ. Обзор решаемых задач и основных методов. Понятия криптографической системы, криптографического алгоритма, криптографического протокола. Обзор нормативной базы в области криптографии. Обзор типовых областей применения криптографических систем, алгоритмов и протоколов. Криптографическая стойкость. Обзор распространенных методов атак.

Технологии распределенного реестра. Блокчейн как тип базы данных, дерево Меркла. Эллиптические кривые над вещественными числами и над конечным полем: получение закрытого и открытого ключей. Трилемма масштабируемости и блокчейн 2.0.

Теория игр

Теорема фон-Неймана о существовании седловой точки в смешанных стратегиях для конечных антагонистических игр двух лиц с нулевой суммой. Теорема о существовании цены игры в случае, когда пространства стратегий игроков компактны, а плата непрерывна.

Равновесие Нэша. Теорема о существовании равновесия в смешанных стратегиях для конечных игр N лиц. Паретовские решения. Ядро игры. Характеристическая функция игры.

Дифференциальные игры двух лиц. Уравнение Гамильтона-Якоби-Айзекса-Беллмана как достаточное условие оптимальности. Принцип минимакса.

Литература

Р. Айзекс Дифференциальные игры. М.: Мир, 1967.

В. М. Алексеев, В. М.Тихомиров, С. В. Фомин Оптимальное управление. М.: Наука, 1979.

С.В.Алешин, Распознавание динамических образов, Изд-во МГУ, М., 1996.

В.А. Артамонов, В.Н. Латышев, Линейная алгебра и выпуклая геометрия, Факториал, 2003.

А.Ахо, Дж. Хопкрофт, Дж. Ульман. Построение и анализ вычислительных алгоритмов. Москва, Мир, 1970 г.

А. Ахо, Дж. Ульман. Теория синтаксического анализа, перевода и компиляции. Том 1: Синтаксический анализ. М.: Мир, 1978 г.

В.А. Галатенко Основы информационной безопасности: курс лекций: учебное пособие. ИНТУИТ. РУ "Интернет-университет Информационных Технологий", 2006.

М. Гери, Д.Джонсон. Вычислительные машины и труднорешаемые задачи. Москва, Мир, 1982 г.

Л.А. Калиниченко, Методы и средства интеграции неоднородных баз данных, М. Наука, 1983.

Л. В. Канторович, .Акилов Функциональный анализ. М.: Изд-во физ.мат.лит. 1977.

Б.У. Керниган, Д.М. Ритчи. Язык программирования С. Вильямс, 2013.

А.Н. Колмогоров, Теория информации и теория алгоритмов, М., 1987
Т Кормен, Ч Лейзерзон, Р.Ривест. Алгоритмы, Построение и анализ. М 1999.

Н. Н. Красовский, А. И. Субботин. Позиционные дифференциальные игры. М.: Наука, 1974.

В.Б. Кудрявцев, С.В. Алешин, А.С. Подколзин. Введение в теорию автоматов. М.Наука. 1985.

В.Б. Кудрявцев, А.Е.Андреев, Э.Э. Гасанов. Теория тестового распознавания. М. ФизМатЛит, 2007.

В.Б. Кудрявцев, Э.Э. Гасанов, А.С. Подколзин. Введение в теорию интеллектуальных систем. М. Изд-во МГУ, 2006.

С.Д. Кузнецов. Базы данных. Академия, Серия: Университетский учебник, 2012.

Р. Лав. Linux. Системное программирование. Питер, 2008.

Ю. Лесковец, А. Раджараман, Дж. Ульман, Анализ больших наборов данных, М.: ДМК Пресс, 2016.

М. Липовач. Изучай Haskell во имя добра. ДМК Пресс, 2012.

Г. Г. Магарил-Ильяев, В. М. Тихомиров, Выпуклый анализ и его приложения. М.: Эдиториал УРСС, 2002.

К. Маннинг, П. Рагхаван, Х. Шютце. Введение в информационный поиск. Вильямс, 2011.

Дж. Мартин. Организация баз данных в вычислительных системах. М.: Мир, 2000.

А.И.Михайлов, А.И.Черный, Р.Э. Гиляревский .Основы информатики. М.: Наука, 1978.

Г. Оуэн. Теория игр. М.: Мир, 1971.

Х Пападимитриу, К. Стайглиц. Комбинаторная оптимизация. Алгоритмы и анализ. М 1985.

И.И. Попов. Информационные ресурсы и системы: реализация, моделирование, управление. М.: ТПК «Альянс», 1996.

Х. Роджерс, Теория рекурсивных функций и эффективная вычислимость, М., 1972.

Р. Рокафеллар. Выпуклый анализ. М.: Мир, 1973.

В.Н. Сачков. Введение в комбинаторные методы дискретной математики. М.2004.

У. Р. Стивенс, Б. Феннер, Э. М. Рудофф. UNIX. Разработка сетевых приложений, 3-е изд. Питер, 2007.

Дж. Ульман, Основа систем баз данных, М. Финансы и статистика, 1983.

С. Фейт. TCP/IP. Архитектура, протоколы, реализация (включая IPv6 и IP Security). Лори, 2009.

Дж.Фон-Нейман, О. Моргенштерн Теория игр и экономическое поведение. .: Наука, 1970.

Дж. Хопкрофт, Р. Мотвани, Дж. Ульман. Введение в теорию автоматов, языков и вычислений. М.: Издательский дом "Вильямс", 2008.

Ю.И. Шемякин Введение в информатику. М.: Финансы и статистика, 1985.

К. Шеннон Работы по теории информации и кибернетике. М., 1963.

D. Allemang, J. Hendler. Semantic web for the working ontologist: effective modeling in RDFS and OWL. – Elsevier, 2011.
G. Antoniou, F. Van Harmelen. A semantic web primer. – MIT press, 2004.

F. Baader, W. Nutt. Basic description logics //Description logic handbook. – 2003. – С. 43-95.

C. Villani. Topics in optimal transportation. Graduate studies in mathematics. Vol. 58, Amer. Math. Soc. Providence, Rhode Island, 2003.