Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение   
высшего образования «Новосибирский национальный исследовательский  
государственный университет» (Новосибирский государственный университет, НГУ)

**Факультет информационных технологий**

СОГЛАСОВАНО

Декан ФИТ НГУ

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ М.М. Лаврентьев

«03» июля 2019 г.

РАБОЧАЯ ПРОГРАММА Дисциплины

|  |
| --- |
| **Интеллектуальная обработка документов и фактов** |

Направление подготовки: 09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

Направленность (профиль): Технология разработки программных систем

Форма обучения: очная

Год обучения: 1, семестр: 2

|  |  |  |
| --- | --- | --- |
| **№** | **Вид деятельности** | **Семестр** |
| **2** |
| **1** | Лекции, час. | 16 |
| **2** | Практические занятия, час. | 32 |
| **3** | Лабораторные занятия, час. |  |
| **4** | Занятий в контактной форме без учета промежуточной аттестации, час, из них | 50 |
| **5** | в электронной форме, час. |  |
| **6** | из них аудиторных занятий, час. | 48 |
| **7** | из них в активной и интерактивной форме, час. | 48 |
| **8** | консультаций, час. | 2 |
| **9** | Самостоятельная работа, час. | 92 |
| **10** | в том числе на выполнение письменных работ, час | 48 |
| **11** | Форма аттестации (экзамен, зачет, дифференцированный зачет), час | Э 2 |
| **12** | Всего зачетных единиц[[1]](#footnote-1) | 4 |

Новосибирск 2019

Рабочая программа дисциплины составлена на основании федерального государственного образовательного стандарта (ФГОС) высшего образования - магистратура по направлению подготовки 09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА.

Федеральный государственный образовательный стандарт (ФГОС) высшего образования по направлению подготовки магистров 09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА введен в действие приказом Минобрнауки 19.09.2017 № 918.

Место дисциплины в структуре учебного плана: Блок 1 Дисциплины (модули), вариативная часть, дисциплина по выбору.

Рабочая программа дисциплины утверждена решением Ученого совета факультета информационных технологий от 02.07.2019, протокол № 75.

Программу разработали:

Профессоркафедры систем информатики ФИТ,

доктор технических наук В.Б. Барахнин

Заведующий кафедрой систем информатики ФИТ,

доктор физико-математических наук М.М. Лаврентьев

Ответственный за образовательную программу:

Заведующий кафедрой систем информатики ФИТ,

доктор физико-математических наук М.М. Лаврентьев

**Аннотация к рабочей программе дисциплины**

**«Интеллектуальная обработка документов и фактов»**

Дисциплина «Интеллектуальная обработка документов и фактов» реализуется в рамках образовательной программы высшего образования – программы магистратуры 09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА, направленность (профиль): Технология разработки программных систем по очной форме обучения на русском языке.

**Место в образовательной программе:** Дисциплина «Интеллектуальная обработка документов и фактов» развивает знания, умения и навыки, сформированные у обучающихся по результатам изучения следующих дисциплин «Современные проблемы информатики и вычислительной техники», «Методологии анализа данных».

Дисциплина «**Интеллектуальная обработка документов и фактов»** является базовой для выполнения работы в рамках практики и выполнением выпускной квалификационной работы.

Дисциплина «**Интеллектуальная обработка документов и фактов»** реализуется во 2 семестре в рамках вариативной части дисциплин (модулей) Блока 1 и является дисциплиной по выбору.

Дисциплина **«Интеллектуальная обработка документов и фактов»** направлена на формирование компетенций:

**Способен осуществлять управление развитием информационной системы организации (ПКС-1) в части следующих индикаторов достижения компетенции:**

ПКС-1.1: Знать принципы организации и функционирования информационных систем;

ПКС-1.2: Уметь анализировать системные проблемы обработки информации на уровне информационной системы;

ПКС-1.3: Уметь работать с информацией в условиях неопределенности, избыточности и недостаточности исходных данных.

**Перечень основных разделов дисциплины:**

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: лекции, практические занятия.

1. Информатика как наука о семантической информации

2. Семиотические методы в информатике.

3. Тезаурусы и онтологии.

4. Документ как основная форма овеществления информации

5. Аналитико-синтетическая переработка документов.

6. Основы фактографического поиска

7. Алгоритмы обработки слабоструктурированных документов.

8. Использование методов машинного обучения для обработки документов.

При освоении дисциплины студенты выполняют следующие виды учебной работы: лекции, практические занятия, консультации, самостоятельная работа. В учебном процессе предусматривается использование активных и интерактивных форм проведения занятий*.*

Самостоятельная работа включает: подготовку к практическим занятиям по разделам дисциплины, подготовку презентаций докладов, подготовку к тестированию, подготовку к экзамену.

Общий объем дисциплины – 4 зачетных единиц (144 часа).

**Правила аттестации по дисциплине.** Текущий контроль по дисциплине «Интеллектуальная обработка документов и фактов» осуществляется на практических занятиях на основании оценки за портфолио (подготовка доклада на одну из заданных тем и прохождение тестирования). По результатам защиты портфолио выставляется оценка «зачтено» или «не зачтено».

Промежуточная аттестация по дисциплине «Интеллектуальная обработка документов и фактов» проводится по завершению периода ее освоения (семестра). Промежуточная аттестация по дисциплине включает 2 этапа:

1) портфолио (подготовка доклада на одну из заданных тем и прохождение тестирования);

2) экзамен.

Оценка «зачтено» за портфолио является необходимым условием для прохождения промежуточной аттестации. Результаты промежуточной (итоговой по дисциплине) аттестации определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Оценки «отлично», «хорошо», «удовлетворительно» означают успешное прохождение промежуточной аттестации.

**Учебно-методическое обеспечение дисциплины.**

Учебно-методический комплекс по дисциплине «Интеллектуальная обработка документов и фактов» в электронной информационно-образовательной среде НГУ:

<https://et.nsu.ru/course/view.php?id=911>

1. **Внешние требования к дисциплине**

Таблица 1.1

|  |  |
| --- | --- |
| ***Компетенция* ПКС-1 Способен осуществлять управление развитием информационной системы организации, *в части следующих индикаторов достижения компетенции:*** | |
| **ПКС-1.1** | Знать принципы организации и функционирования информационных систем |
| **ПКС-1.2** | Уметь анализировать системные проблемы обработки информации на уровне информационной системы |
| **ПКС-1.3** | Уметь работать с информацией в условиях неопределенности, избыточности и недостаточности исходных данных |

1. **Требования к результатам освоения дисциплины**

Таблица 2.1

|  |  |  |  |
| --- | --- | --- | --- |
| **Результаты изучения дисциплины по уровням освоения (иметь представление, знать, уметь, владеть)** | **Формы организации занятий** | | |
| **Лекции** | **Практики / семинары** | **Самостоятельная работа** |
| ПКС-1.1 Знать принципы организации и функционирования информационных систем | | | |
| 1. Знать принципы организации и функционирования систем интеллектуальной обработки документов и фактов. | + | + | + |
| ПКС-1.2 Уметь анализировать системные проблемы обработки информации на уровне информационной системы | | | |
| 2. Уметь проводить сравнительный анализ вариантов решения задач аналитико-синтетической переработки документов и обработки фактов. | + | + | + |
| ПКС-1.3 Уметь работать с информацией в условиях неопределенности, избыточности и недостаточности исходных данных | | | |
| 3. Уметь применять типовые алгоритмы аналитико-синтетической переработки документов. и обработки фактов. | + | + | + |

**3. Содержание и структура учебной дисциплины**

Таблица 3.1

|  |  |  |  |
| --- | --- | --- | --- |
| **Темы лекций** | **Активные формы, час.**  **(входит в общее кол-во часов)** | **Часы** | **Ссылки на результаты обучения** |
| **Семестр: 2** | | | |
| 1. Информатика как наука о семантической информации | 2 | 2 | 1 |
| 1. Семиотические методы в информатике. | 2 | 2 | 1 |
| 1. Тезаурусы и онтологии. | 2 | 2 | 1 |
| 1. Документ как основная форма овеществления информации | 2 | 2 | 1 |
| 1. Аналитико-синтетическая переработка документов. | 2 | 2 | 1,2 |
| 1. Основы фактографического поиска. | 2 | 2 | 1,2 |
| 1. Алгоритмы обработки слабоструктурированных документов. | 2 | 2 | 1,3 |
| 1. Использование методов машинного обучения для обработки документов. | 2 | 2 | 1,3 |
| **Итого:** | **16** | **16** |  |

Таблица 3.2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Темы практических занятий** | **Активные формы, час.**  **(входит в общее кол-во часов)** | | **Часы** | **Ссылки на результаты обучения** | | **Учебная деятельность** |
| **Семестр: 2** | | | | | | |
| Тема 1. Системы обработки семантической информации: история и современность. | 6 | 6 | | | 1 | Обучающиеся изучают и анализируют тенденции развития систем обработки семантической информации, выступают с докладами. |
| Тема 2. Тезаурусы и онтологии. Сравнительный обзор классификаторов. | 6 | 6 | | | 1 | Обучающиеся изучают и анализируют технологии создания тезаурусов и онтологий, а также классификаторы, включая вопросы обоснования их выбора, выступают с докладами. |
| Тема 3. Форматы представления документов (текстовых, графических, аудио, видео…) и их особенности. Алгоритмы поиска документов нетекстовых форматов. | 6 | 6 | | | 1,2,3 | Обучающиеся изучают и анализируют форматы представления документов, а также вопросы обоснования выбора алгоритмов обработки документов в зависимости от формата их представления, выступают с докладами. |
| Тема 4. Индексация документов. Метаданные и обработка электронных ресурсов. | 4 | 4 | | | 1,2,3 | Обучающиеся изучают и анализируют вопросы метаописания документов, включая алгоритмы индексации, выступают с докладами |
| Тема 5. Особенности онтологий для фактографических систем. Автоматизированное извлечение фактов из документов. О взаимодействии фактографических систем с пользователями. | 4 | 4 | | | 1,2,3 | Обучающиеся изучают и анализируют вопросы создания фактографический систем, включая сравнительный анализ алгоритмов обработки фактов, выступают с докладами. |
| Тема 6. Кластеризация текстовых документов | 4 | 4 | | | 1,2,3 | Обучающиеся изучают и анализируют алгоритмы кластеризации текстовых документов, включая вопросы обоснования их выбора, выступают с докладами. |
| Тема 7. Использование методов машинного обучения для обработки документов | 2 | 2 | | | 1,2,3 | Обучающиеся изучают и анализируют методы машинного обучения для обработки документов, выступают с докладами. |
| **Итого:** | **32** | **32** | | |  |  |

1. **Самостоятельная работа студентов**

Таблица 4.1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **№** | **Виды самостоятельной работы** | | | **Ссылки на результаты обучения** | | **Часы на выполнение** | | | **Часы на консультации** |
| **Семестр: 2** | | | | | | | | | | |
| 1 | | Изучение разделов дисциплины по учебной литературе, в том числе вопросов, не освещаемых на лекциях | | 1, 2, 3 | | 20 |  | | | |
| Изучение предлагаемых теоретических разделов в соответствии с настоящей Программой. Учебно-методические материалы по дисциплине «Интеллектуальная обработка документов и фактов» выложены на странице курса в сети Интернет | | | | | | | | |
| 2 | | Подготовка к практическим занятиям, к текущему контролю знаний | | 1, 2, 3 | | 46 |  | | | |
| Подготовка доклада, выполнение индивидуального проекта | | | | | | | | |
| 3 | | Подготовка к экзамену | | 1, 2, 3 | | 26 | 2 | | | |
| Повторение теоретического материала по вопросам, совпадающим с темами лекций | | | | | | | | |
|  | | Итого |  | | 92 | | | 2 | | |

1. **Образовательные технологии**

В ходе реализации учебного процесса по дисциплине проводятся лекционные и практические занятия. Темы, рассматриваемые на лекциях и изучаемые самостоятельно, закрепляются на практических занятиях, по вопросам, вызывающим затруднения, проводятся консультации на практических занятиях. Применяются такие формы проведения практических занятий, как обсуждение и защита результатов работы, а также используются следующие интерактивные формы обучения (таблица 5.1).

Таблица 5.1

|  |  |  |
| --- | --- | --- |
| **1** | Технологии проблемного обучения | ПКС-1.1, 1.2, 1.3 |
| **Формируемые умения:** 2. Уметь проводить сравнительный анализ вариантов решения задач аналитико-синтетической переработки документов и обработки фактов. | | |
| **Краткое описание применения:** Постановка под руководством преподавателя проблемных задач и активная самостоятельная деятельность обучающихся по их разрешению, сопровождающаяся обсуждением результатов. | | |
| **2** | Портфолио | ПКС-1.1, 1.2, 1.3 |
| **Формируемые умения:** 3. Уметь применять типовые алгоритмы аналитико-синтетической переработки документов и обработки фактов. | | |
| **Краткое описание применения:** студенты ведут портфолио (коллекцию работ), которое является основой для проведения аттестации по дисциплине. | | |

Для организации и контроля самостоятельной работы студентов, а также проведения консультаций применяются информационно-коммуникационные технологии (таблица 5.2).

Таблица 5.2

|  |  |
| --- | --- |
| Информирование | https://et.nsu.ru/course/view.php?id=911 |
| Консультирование | https://et.nsu.ru/course/view.php?id=911 |
| Контроль | https://et.nsu.ru/course/view.php?id=911 |
| Размещение учебных материалов | - |

1. **Правила аттестации студентов по учебной дисциплине**

По дисциплине **«Интеллектуальная обработка документов и фактов»** проводится текущая и промежуточная аттестация (итоговая по дисциплине).

**Текущая аттестация** по дисциплине **«Интеллектуальная обработка документов и фактов»** осуществляется на практических занятиях и заключается в презентации и защите докладов по каждой теме практических занятий. В ходе обучения каждый студент должен подготовить презентации докладов по каждому разделу самостоятельной работы и публично выступить с ними, защищая полученные результаты в ходе обсуждения и дискуссии. По результатам текущей аттестации выставляется оценка «зачтено» или «не зачтено». Оценка «зачтено» по результатам защиты докладов является одним из условий успешного прохождения промежуточной аттестации.

Для получения оценки «зачтено» презентация и доклад на каждую тему, соответствующую разделам дисциплины в каждом семестре, должна быть выполнена и защищена в полном соответствии с предъявляемыми требованиями.

**Промежуточная аттестация** по дисциплине **«Интеллектуальная обработка документов и фактов»** проводится по завершению периода ее освоения (семестра). Промежуточная аттестация по дисциплине включает 2 этапа:

1) портфолио (подготовка доклада на одну из заданных тем и прохождение тестирования);

2) экзамен.

Оценка «зачтено» за портфолио является необходимым условием для прохождения промежуточной аттестации. Результаты промежуточной аттестации определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Оценки «отлично», «хорошо», «удовлетворительно» означают успешное прохождение промежуточной аттестации.

В таблице 6.1 представлено соответствие форм аттестации заявляемым требованиям к результатам освоения дисциплины.

Таблица 6.1

|  |  |  |  |
| --- | --- | --- | --- |
| **Коды компетенций ФГОС** | **Результаты обучения** | **Формы аттестации** | |
| **1 этап - портфолио** | **2 этап - экзамен** |
| **ПКС-1** | **ПКС-1.1** Знать принципы организации и функционирования информационных систем | **+** | **+** |
| **ПКС-1.2** Уметь анализировать системные проблемы обработки информации на уровне информационной системы | **+** | **+** |
|  | **ПКС-1.3** Уметь работать с информацией в условиях неопределенности, избыточности и недостаточности исходных данных | **+** | **+** |

Требования к структуре и содержанию портфолио, оценочные средства, а также критерии оценки сформированности компетенций и освоения дисциплины в целом, представлены в Фонде оценочных средств, являющемся приложением 1 к настоящей рабочей программе дисциплины.

1. **Литература**
2. Шокин, Юрий Иванович. Проблемы поиска информации / Ю.И. Шокин, А.М. Федотов, В.Б. Барахнин; отв. ред. О.Л. Жижимов; Рос. акад. наук, Сиб. отд-ние, Ин-т вычисл. технологий. Новосибирск: Наука, 2010. 197 с.: ил., табл., [1] л. портр.; 22 см. ISBN 978-5-02-018969-0. (15 экз.)
3. Федотов, Анатолий Михайлович. Информационные системы модели и технологии : учеб.пособие / А.М. Федотов, О.А. Федотова, М.А.Самбетбаева / отв.ред.В.Б.Барахнин; Новосиб. гос. ун-т. Новосибирск: ИПЦ НГУ, 2019. 264 с. ISBN 978-5-4437-0857-7.

<https://e-lib.nsu.ru/reader/bookView.html?params=UmVzb3VyY2UtNTM2MA/cGFnZTAwMDAw>

1. Батура Татьяна Викторовна. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учебное пособие : [для студентов и аспирантов ФИТ, ММФ и ГФ(отделение фундаментальной и прикладной лингвистики) НГУ] / Т.В. Батура ; М-во образования и науки Рос. Федерации, Новосиб. гос. ун-т, Фак. информ. технологий, Каф. систем информатики.— Новосибирск : Редакционно-издательский центр НГУ, 2016 .— 165 с. : ил.; 20 см. — Библиогр. в конце глав .— ISBN 978-5-4437-0548-4

<https://e-lib.nsu.ru/reader/bookView.html?params=UmVzb3VyY2UtMTU4Mw/cGFnZTAwMQ>

*Интернет-ресурсы*

Таблица 7.1

|  |  |  |
| --- | --- | --- |
| № п/п | Наименование Интернет-ресурса | Краткое описание |
| 1 | Журнал «Вестник НГУ. Серия: Информационные технологии» [Электронный ресурс]. – Режим доступа: <https://journals.nsu.ru/jit/>. – Загл. с экрана | Полнотекстовые электронные копии статей в области вычислительный методов (с 2006 года). |
| 2 | <https://e-lib.nsu.ru/reader/bookView.html?params=UmVzb3VyY2UtMTkwOA/cGFnZTAwMQ> | Барахнин В.Б., Федотов А.М. Построение модели фактографического поиска // Вестник НГУ. Серия: Информационные технологии. – 2013. – Т. 11. – Вып. 4 – С. 16-27. |

**8. Учебно-методическое и программное обеспечение дисциплины**

**8.1. Учебно-методическое обеспечение**

Барахнин В.Б. Интеллектуальная обработка документов и фактов [Электронный ресурс]: электронный учебно-методический комплекс / В.Б. Барахнин; Новосиб. гос. ун-т. - Новосибирск, [2018]. - Режим доступа: <https://et.nsu.ru/course/view.php?id=911>. - Загл. с экрана.

**8.2. Программное обеспечение**

Для обеспечения реализации дисциплины используется стандартный комплект программного обеспечения (ПО), включающий регулярно обновляемое лицензионное ПО Windows и MS Office.

Перечень специализированного программного обеспечения для изучения дисциплины представлен в таблице 8.1.

Специализированное программное обеспечение Таблица 8.1

|  |  |  |
| --- | --- | --- |
| **№** | **Наименование ПО** | **Назначение** |
| 1 | Microsoft Visual Studio Professional 2019 | Среда разработки приложений |
| 2 | Eclipse 2019 | Среда разработки приложений |

**9. Профессиональные базы данных и информационные справочные системы**

1. Полнотекстовые электронные ресурсы Freedom Collection издательства Elsevier (Нидерланды) (2 предметные коллекции – Computer Science, Mathematics)
2. БД Scopus (Elsevier)

**10. Материально-техническое обеспечение**

Таблица 10.1

|  |  |  |
| --- | --- | --- |
| **№** | **Наименование** | **Назначение** |
| 1 | Презентационное оборудование (мультимедиа-проектор, экран, компьютер для управления) | Для проведения лекционных занятий |
| 2 | Компьютерный класс (с выходом в Internet) | Для организации практических занятий и самостоятельной работы обучающихся |

Материально-техническое обеспечение образовательного процесса по дисциплине для обучающихся из числа лиц с ограниченными возможностями здоровья осуществляется согласно «Порядку организации и осуществления образовательной деятельности по образовательным программам для инвалидов и лиц с ограниченными возможностями здоровья в Новосибирском государственном университете».

**Лист актуализации рабочей программы дисциплины  
«Интеллектуальная обработка документов и фактов»**

|  |  |  |  |
| --- | --- | --- | --- |
| № | Характеристика внесенных изменений (с указанием пунктов документа) | Дата и №  протокола Ученого совета ФИТ | Подпись  ответственного |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. С учетом выделенных часов на промежуточную аттестацию [↑](#footnote-ref-1)