



**Федеральное государственное автономное образовательное учреждение
высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет гуманитарных наук
Школа лингвистики

**Рабочая программа дисциплины
Автоматическая обработка естественного языка**

для образовательной программы "Фундаментальная и компьютерная лингвистика"
направления подготовки 45.03.03. Фундаментальная и прикладная лингвистика
уровень бакалавр

Разработчик(и) программы
Толдова С.Ю., к.ф.н., stoldova@hse.ru

Одобрена на заседании Школы лингвистики ФГН «30» мая 2016 г.
Руководитель Школы лингвистики Е.В. Рахилина

Рекомендована Академическим советом образовательной программы
«01» июня 2016 г., № протокола 10

Утверждена «01»июня 2016 г.
Академический руководитель образовательной программы
Ю.А. Ландер

___Москва___, 2016_

*Настоящая программа не может быть использована другими подразделениями университета
и другими вузами без разрешения подразделения-разработчика программы.*



1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину Автоматическая обработка естественного языка, учебных ассистентов и студентов направления подготовки направления, обучающихся по образовательной программе Фундаментальная и прикладная лингвистика.

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ ОС Образовательной программой направления подготовки 45.03.03. Фундаментальная и прикладная лингвистика "Фундаментальная и прикладная лингвистика".
- Объединенным учебным планом университета по образовательной программе Фундаментальная и прикладная лингвистика, утвержденным в 2016 г.

2 Цели освоения дисциплины

Целями освоения дисциплины Автоматическая обработка естественного языка являются знакомство с основными проблемами в области компьютерной лингвистики, базовыми алгоритмами, математическими методами моделирования языковых феноменов, основными инструментами и технологиями в области автоматической обработки естественного языка, умение представлять в алгоритмическом виде процессы анализа и синтеза текста.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

[Компетенции для программы учебной дисциплины берутся из стандартов: ФГОС/ ОС НИУ ВШЭ для соответствующего уровня и направления подготовки и из числа закрепленных за дисциплиной в матрице компетенций образовательной программы]

Уровни формирования компетенций:

РБ — ресурсная база, в основном теоретические и предметные основы (знания, умения);

СД – способы деятельности, составляющие практическое ядро данной компетенции;

МЦ – мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать

В результате освоения дисциплины студент осваивает компетенции:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен планировать научно-исследовательскую деятельность, проводить самостоятельные	ПК-1	РБ, МЦ, СД	Дает определения основным понятиям автоматической обработки текста, воспроизводит базовые алгоритмы, используемые в автоматической обработке текста, использует основные	- чтение специальной литературы -выполнение самостоятельных заданий	Тест, проверка дз, сдача проекта



Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
исследования и получать новые научные результаты в области профессиональной деятельности			пакеты морфологической обработки текста, демонстрирует знание базовых алгоритмов, владеет, использует современные методы тестирования качества, применяет современные подходы к решению задач в области компьютерной лингвистики, интерпретирует результаты базовых алгоритмов] ¹	- анализ полученных данных	
способен проводить формализацию лингвистических знаний, анализ и синтез лингвистических структур, количественный анализ лингвистических данных с использованием математических знаний и методов	ПК-2	РБ, МЦ, СД	дает определения основным формальным системам, используемым при разработке алгоритмов по морфологическому анализу текста, дает определения основным этапам синтаксического и морфологического анализа, воспроизводит различные алгоритмы синтаксического разбора, распознает основные типы ошибок автоматического и морфологического автоматического анализа, применяет количественные подходы к обработке текста и выделению ключевых слов в тексте, демонстрирует знакомство с базовыми методами машинного обучения, применяемого для семантического анализа (вычисление коллокаций, выделение ключевых слов, разрешение семантической неоднозначности)		
способен участвовать в создании представительных текстовых массивов, корпусов текстов, корпусов звучащей речи, мультимодальных корпусов, лингвистических и	ПК-11	РБ, СД	знает основные лингвистические ресурсы, владеет методами разметки корпусов и составлением частотных списков, владеет базовыми статистическими методами подготовки метарилла к созданию лексикографических ресурсов	практические занятия по созданию языковых ресурсов и лингвистических компонентов обзор и рефераты существующ	Практические занятия, выполнение упражнений, работа с ресурсами на

¹ В шаблоне дан неполный перечень глаголов – подсказок. Возможно использование и других формулировок дескрипторов.



Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
социолингвистических баз данных и пользоваться этими ресурсами				их разработок	практических занятий
способен проектировать системы анализа и синтеза естественного языка, анализа и синтеза мультимодальных языковых систем, в том числе лингвистических компонентов интеллектуальных и информационных электронных систем	ПК-12	РБ, СД	Знает наиболее известные доступные для свободного использования компоненты автоматического анализа, умеет использовать соответствующие модули в различных приложениях, в том числе синтаксические и морфологические парсеры, создавать модули первичной обработки текста	Практические задания, выполнение проекта в группах	

4 Место дисциплины в структуре образовательной программы

Для ОС НИУ НИУ:

Настоящая дисциплина относится к профессиональному блоку дисциплин «Компьютерная лингвистика».

Для специализаций Фундаментальная и компьютерная лингвистика профиля компьютерная лингвистика настоящая дисциплина является базовой.

Изучение данной дисциплины базируется на следующих дисциплинах:

- Введение в лингвистику (первый и второй курс) программы подготовки бакалавра
- Теория языка
- Линейная алгебра и математический анализ программы подготовки бакалавра
- Программирование и компьютерные инструменты лингвистического исследования программы подготовки бакалавра
- Дискретная математика программы подготовки бакалавра
- Программирование и компьютерные инструменты лингвистического исследования программы подготовки бакалавра
- Иностранный язык
- Автоматическая обработка естественного языка (3-ий курс)

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- владеть базовыми представлениями о грамматических категориях и анализе языковых единиц;
- владеть базовыми знаниями в области теории алгоритмов и основ математики
- владеть базовыми знаниями в области теории вероятностей и статистики;



- уметь читать научные работы и технические описания на английском языке;
- запускать готовые пакеты для морфологической и синтаксической обработки текстов.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении дисциплин:

- информационный поиск и извлечение данных
- программирование для лингвистов
- подготовка и защита выпускной квалификационной работы

5 Тематический план учебной дисциплины

[Тематический план отражает содержание дисциплины (перечень разделов), структурированное по видам учебных занятий с указанием их объемов в соответствии с ОУП]

№	Название раздела	Всего часов	Аудиторные часы				Самостоятельная работа
			Лекции	Семинары	Практические занятия	Другие виды работы ²	
	Введение. Обзор приложений с использованием лингвистических модулей		2				
1	Квантитативные характеристики слов и методы автоматического семантического анализа	42	6	6			30
2	Кластеризация, классификация текстов, методы машинного обучения	46	8	8			32
3	Задачи автоматического анализа контента	52	4	6			34
4	Подготовка к контрольной и тестам	16					8
6	Итого	144	20	20			104

6 Формы контроля знаний студентов

Тип контроля	Форма контроля				Параметры **
		2	3	4	
Текущий	Контрольная работа		*		письменная работа 120 минут
	Домашнее задание 1		*		Отчет по выполнению задания, требования к отчету регламентируются описанием требований к отчету по проекту
	Домашнее задание 2		*		Отчет по выполнению задания, требования к отчету регламентируются описанием требований к отчету по проекту
	Проект по курсу		*		Отчет по выполнению задания, требования к отчету регламентируются описанием требований к отчету по проекту
	Другие				

² Указать другие виды аудиторной работы студентов, если они применяются при изучении данной дисциплины.



	формы (указать)			
Итогов ый	Экзамен		*	Устный. В форме ответов на вопросы и прак- тических заданий.

** В графе Параметры указывается регламент (свод постоянных или временных правил, регулирующих внутреннюю организацию и формы деятельности) проведения контроля (заполняется для каждого контроля соответственно): формат работы (письменная, устная, тест, тест в компьютерной программе и другое), время, отведенное на аудиторские работы, количество дней проведения контроля, количество дней оценки результатов контроля (только для итогового контроля), объем письменных работ для домашних работ, сроки сдачи письменных работ (число), время на самостоятельную подготовку письменных работ и другая информация, носящая регламентирующий характер.]

7 Критерии оценки знаний, навыков

Данный курс в качестве текущего контроля предполагает выполнение практических заданий на семинарах и практических занятиях, а также выполнение трех домашних проектных заданий, результаты которых представляются студентами (в группах) в виде проектов на языке Python, а также технических отчетов по результатам выполнения проектов. Кроме того, на практических занятиях осуществляется постоянный текущий контроль в форме заданий, тестов, а также обсуждения текущих домашних упражнений. В рамках курса планируется выполнение группового проекта и его защита. Проект представляет собой описание проекта анализа контента, включающего модули анализа текста. Он должен содержать описание задачи, анализ аналогов и подходов, описание обучающей и тестовой коллекции, описание структуры лингвистических модулей и подробный разбор одного из модулей.

Задания и проверка производится через LMS. Задания выполняются в формате проектов в LMS. Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

8 Содержание дисциплины

Раздел представляется в удобной форме (список, таблица). Изложение строится по разделам и темам. Содержание темы может распределяться по лекционным и практическим занятиям.

	Лекции	Практи- ческие занятия	Самосто- ятельная работа
1. Введение	2		
<u>Раздел 1. Квантитативные характеристики слов и использова- ние их в автоматической обработке текста. Автоматический семантический анализ</u>			
Методы выделения тематически значимых слов в тексте. Ключе- вые слова. Tf.idf и другие меры тематической значимости	2	2	4
Методы выделения устойчивых словосочетаний	2	2	4
Методы разрешения семантической неоднозначности	2	2	6
Проект по применению квантитативных методов в лексико- графии			16
	6	6	30
<u>Раздел 2. Первичная обработка текста.</u>			
Введение в машинное обучение	2	2	2
Методы кластеризации текстов	2	2	4
Методы классификации текстов	2	2	4
Использование пакета WEKA для автоматической классифика- ции и кластеризации	2	2	6



Подготовка домашнего задания по разделу (проект по классификации / кластеризации текстов)			16
	6	8	32
Раздел 3. Задачи и проекты компьютерной лингвистики			
Обзор задач и систем компьютерной лингвистики	4		
Постановка задач для выполнения курсового проекта: разработка ТЗ для создания системы обработки текста (синтеза текста, машинного перевода и т.п.), использующей модули автоматической обработки текста ³ . Обсуждение тем проектов.		2	10
Обсуждение постановки задачи для каждого из выбранных студентами проектов. Презентации и тестирование существующих систем, использующих модули АОР, аналогичных системам, выбранным для проекта		2	10
Взаимное рецензирование и обсуждение проекта: презентации курсовых проектов. Обсуждение результатов анализа проекта оппонировавшей группы ⁴		2	4
Подготовка окончательной версии проекта			14
Рецензирование проекта другой группы			4
	4	6	42

9 Образовательные технологии

[Основой для знакомства с методологией анализа текстов и овладения практическими навыками автоматического анализа текста с помощью программных средств в рамках курса служит работа над двумя учебными проектами по анализу текстовой коллекции, а также курсовой проект, представляющий собой составление аналитической записки и проектного задания на создание продукта, включающего компонент автоматической обработки текста. В начале курса студентам предлагается обзор примеров таких проектов (например, автоматический генератор стихов, диалоговая система подбора некоторого продукта и т.д.). На протяжении курса студенты работают над проектом в группах (3-4 человека): выбором темы проекта, исследованием аналогов, тестированием существующих модулей лингвистического анализа, которые могли бы быть использованы при реализации проекта. В задачи проекта входит описание проекта разработки лингвистического модуля. По результатам анализа группы пишут отчет по проекту, готовят презентации. В систему оценивания работы по проектам также входит взаимное рецензирование проектов других групп.

В качестве простых упражнений рекомендуется также упражнения на пошаговое применение обсуждаемых в курсе алгоритмов вручную. Для освоения материала на практических занятиях используются задания, предназначенные как для индивидуального решения задач по обработке текста, так и для коллективного обсуждения стратегии решения той или иной задачи. Особое внимание уделяется организации самостоятельной работы студентов с программным обеспечением, рассматриваемым в рамках курса.

³ Курс делится на группы по 3 человека. Каждая группа выбирает одну из тем по актуальным направлениям и готовит курсовой проект: проект ТЗ на создание продукта, содержащего лингвистический компонент и использующего технологии автоматической обработки текста с проработкой лингвистического модуля системы.

⁴ Проекты подвергаются взаимному рецензированию: каждой группе назначается группа-оппонент, задача которой ознакомиться с проектом, проанализировать, написать рецензию, изложить критические замечания и задать необходимые вопросы



10 Оценочные средства для текущего контроля и аттестации студента

10.1 Оценочные средства для оценки качества освоения дисциплины в ходе текущего контроля

Примерные вопросы/ задания для домашних работ:

1. Вычислите $tf.idf$ для выбранного Вами текста и выбранного корпуса текстов
2. Протестируйте систему выделения ключевых слов. Вычислите точность и полноту
3. Разметьте в корпусе текстов глаголы по значениям
4. Дано множество контекстов некоторого многозначного слова. Сгруппируйте контексты по семантической близости. Разбейте множество контекстов на группы в соответствии с конкретным значением слова.
5. Предложите вариант проекта системы, использующей модули автоматической обработки текста. Обоснуйте актуальность и новизну такой системы
6. Выполните анализ аналогов системы
7. Составьте предварительное описание проекта
8. Проведите рецензирование проекта другой группы

10.2 Вопросы для оценки качества освоения дисциплины

1. Какие методы выделения ключевых слов вы знаете
2. Какие методы выделения коллокаций Вы знаете; каковы параметры выделения устойчивых словосочетаний
3. Какие методы разрешения семантической неоднозначности, основанные на базах знаний, Вы знаете?
4. Какие статистические методы применяются в задачах разрешения семантической неоднозначности
5. Назовите основные методы кластеризации. Как эти методы применяются к задаче кластеризации текстов
6. Какие методы классификации применяются в задачах рубрикации текстов
7. Каковы задачи извлечения именованных сущностей? Какие типы омонимии необходимо разрешать в задачах автоматического извлечения сущностей
8. В чем заключается задача и извлечения фактов и отношений? Какие два базовых подхода используются в решении данной задачи? Приведите примеры систем.
9. Перечислите задачи извлечения мнений и анализа тональности.
10. Назовите основные классы лексем и конструкций, которые необходимо учитывать в автоматическом анализе тональности и извлечении мнений.

10.3 Примеры заданий промежуточного /итогового контроля

По желанию автора программы, приводятся примеры билетов с вопросами и задачами, заданий для зачета или экзамена, тренировочные тесты по дисциплине.

Примеры домашних заданий:

Основной задачей курса является написание курсового проекта, который представляет собой проект ТЗ по разработке системы, решающих одну из задач в рамках одного из актуальных направлений компьютерной лингвистики. Студенты разбиваются на группы по 3-4 человека. Каждая группа исследует материалы по актуальным направлениям компьютерной лингвистике, обсуждаемых на соответствующих порталах по компьютерной лингвистике (например, создание систем извлечения фактов из новостных сообщений, мультязычного тезауруса по конкретной предметной области и т.п.), выбирает одно из направлений. Написание проекта разбивается на 6 подзадач, каждая из которых обсуждается на семинарах:

- Задача 1. Анализ порталов по компьютерной лингвистике, указанных преподавателем, выбор темы – типа системы;
- Задача 2. Подготовка драфта проекта; списка вопросов, которые должны быть уточнены на семинаре; темы лекции, которая необходима для написания проекта;



- Задача 3. Тестирование существующих аналогичных систем, подготовка к обсуждению и тестированию аналогов на семинаре
- Задача 4. Разработка проекта ТЗ
- Задача 5. Анализ одного из проектов коллег. Рецензирование проекта
- Задача 6. Внесение изменений, уточнений, исправлений по результатам рецензирования и обсуждения проекта на семинаре. Подготовка окончательного варианта проекта.

Примерный план проекта:

1. Название проекта, разработчики проекта, вклад каждого разработчика
2. Назначение объекта проектирования: какие задачи будет решать система?
3. Области применения: где и кем такая система может применяться, оценить востребованность системы
4. Уникальность системы: есть ли аналоги в России и на западе.
5. Общее описание работы системы. Алгоритм работы.
6. Детальное описание лингвистических модулей.
7. Оценка времени разработки и необходимого количества разработчиков
8. Методика проведения тестирования системы

11 Порядок формирования оценок по дисциплине

Преподаватель оценивает работу студентов на практических занятиях: активность студентов в дискуссиях, правильность выполнения заданий (на основании отчетов по практическим заданиям), выполнение мини-тестов по теории. Оценки за работу на практических занятиях преподаватель выставляет в рабочую ведомость. Оценка по 10-ти балльной шкале за работу на практических занятиях определяется перед промежуточным или итоговым контролем - *O_{аудиторная}*.

Преподаватель оценивает самостоятельную работу студентов: правильность выполнения домашних работ, задания для которых выдаются на практических занятиях (*имеются ввиду домашние работы, которые не включаются в ОУП, это не форма текущего контроля "Домашнее задание"*), полнота освещения темы, которую студент готовит для выступления с докладом на занятии-дискуссии. Оценки за самостоятельную работу студента преподаватель выставляет в рабочую ведомость. Оценка по 10-ти балльной шкале за самостоятельную работу определяется перед промежуточным или завершающим контролем - *O_{сам}*.

Накопленная оценка по дисциплине рассчитывается по формуле:

$$O_{\text{накопленная}} = 0,7 * O_{\text{текущий}} + 0,1 * O_{\text{ауд}} + 0,2 * O_{\text{сам.работа}}$$

Накопленная оценка за текущий контроль учитывает результаты студента по текущему контролю следующим образом:

$$O_{\text{текущий}} = 0,3 \cdot O_{\text{контр}} + 0,7 \cdot O_{\text{дз}};$$

Способ округления накопленной оценки текущего контроля: – арифметический.

В диплом выставляется результирующая оценка по учебной дисциплине.

$$O_{\text{результ}} = 0,8 * O_{\text{накопл}} + 0,2 * O_{\text{экз}}$$

Способ округления результирующей оценки по учебной дисциплине: [арифметический].

Примеры расчета оценки приведены в Приложении 1.

12 Учебно-методическое и информационное обеспечение дисциплины

12.1 Базовый учебник

[*Jurafsky, D., Martin, J. H.* (2000) *Speech and language processing*. NJ: Prentice Hall, 2000..
http://www.deepsky.com/~merovech/voynich/voynich_manchu_reference_materials/PDFs/jurafsky_martin.pdf



12.2 Основная литература

Jurafsky, Daniel, and James H. Martin. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. 2nd edition. Prentice-Hall. Ch.20, Ch. 22

Manning C.D., Schütze H. *Foundations of Statistical Natural Language Processing* - Cambridge, Massachusetts: The MIT Press, 1999. Ch. 5. Collocations.

[Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце](#) Введение в информационный поиск, М. Вильямс, 201

12.3 Дополнительная литература

Salton G, Wong A, Yang C. A Vector Space Model for Automatic Indexing. // *Communications of the ACM*, 18(11):613-620, 1975.

Manning C.D., Schütze H. *Foundations of Statistical Natural Language Processing* - Cambridge, Massachusetts: The MIT Press, 1999.

Браславский П., Соколов Е. Сравнение пяти методов извлечения терминов произвольной длины // *Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог»* (Бекасово, 4–8 июня 2008 г.). Вып. 7 (14).– М.: РГГУ, 2008. - С. 67-74.

Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце. Введение в информационный поиск - 2011 г. С. 257 – 294.

Литература к проектам

Прикладная лингвистика. // Статья в энциклопедии «Фонд знаний «Ломоносов»».

URL: <http://www.lomonosov-fund.ru/enc/ru/encyclopedia:01206:article>.

Литература для докладов (презентации на коллоквиуме):

Генерация текстов

Соколова Е.Г., Болдасов М. Автоматическая генерация текстов на ЕЯ (портрет направления) <http://www.dialog-21.ru/Archive/2004/Sokolova.htm>

Ehud Reiter. Has a Consensus NL Generation Structure Appeared, and is it Psycholinguistically Plausible? — 7th International Generation Workshop (Kennebunkport, Maine). URL: <http://www.aclweb.org/anthology/W/W94/W94-0319.pdf>

Машинный перевод

Лекция Л.Л.Иомдина “Машинный перевод: успехи, неудачи, надежды”. Лекторий Политехнического музея. Видео. Доступно с URL

<http://rutube.ru/video/828268c50a49b876a6f4676b839fa745/> дата обращения 20.01.2015)

3. Brown P. F. et al. The mathematics of statistical machine translation: Parameter estimation // *Computational linguistics*. – 1993. – Т. 19. – №. 2. – С. 263-311. <http://acl.ldc.upenn.edu/J/J93/J93-2003.pdf>

Text mining (классификация, кластеризация, реферирование). На примере анализа новостного потока

Кондратьев М. Е. Анализ методов кластеризации новостного потока // *Тр. Восьмой Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL'2006)*.—Ярославль. – 2006. – С. 108-114. rcdl.ru/doc/2006/paper_92_v1.pdf

Распознавание речи

Speech recognition. http://en.wikipedia.org/wiki/Speech_recognition

В. Н. Сорокин, *Синтез речи*. М., 1992.,

D. Childers, *Speech Processing*, John Wiley and Sons, Inc., все издания, начиная с 1999

Диалоговые системы. Говорящие роботы

Filipe P. P., Morgado L., Mamede N. J. An Adaptive Domain Knowledge Manager for Dialogue Systems // *ICEIS* (5). – 2007. – С. 45-52. <http://www.inesc->



id.pt/pt/indicadores/Ficheiros/3983.pdf

Bermúdez M. G., Vila M. G. Dialogue Management for multilingual communication through different channels.

Извлечение информации из текста: извлечение именованных сущностей, отношений и фактов

Information Extraction http://en.wikipedia.org/wiki/Information_extraction

Котельников Д. С., Лукашевич Н. В. Итерационное извлечение шаблонов описания событий по новостным кластерам. <http://ceur-ws.org/Vol-934/paper45.pdf>

Nadeau D. and Sekine S. A survey of named entity recognition and classification, *Linguisticae Investigationes*, Amsterdam, Netherlands: John Benjamins Publishing Company, 1: Vol. 30. pp. 3-26.

Brykina M., Toldova S.Yu., Faynveyts A. V. Dictionary-based ambiguity resolution in Russian named-entities recognition. A case study. P. 163-177. *Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог» 2013. Вып. 12(19). М.: РГГУ, 2013.* URL: <http://www.dialog-21.ru/digests/dialog2013/materials/pdf/BrykinaMM.pdf>

Пазельская А. Г., Соловьев А. Н. Метод определения эмоций в текстах на русском языке //ЗАО «Ай-Теко», Москва, Россия. URL: <http://www.dialog-21.ru/digests/dialog2011/materials/ru/pdf/50.pdf>

12.4 Справочники, словари, энциклопедии

12.5 Программные средства

Для успешного освоения дисциплины, студент использует следующие программные средства:

- Программные среды TOMITA Parser, NLTK
- Python
- WEKA

13 Материально-техническое обеспечение дисциплины

Для проведения практических занятий необходимы компьютерные классы, для проведения лекций проектор, экран



Примеры подсчета оценки за дисциплину в различных случаях

Накопленная оценка за текущий контроль учитывает результаты студента по текущему контролю следующим образом:

$$O_{\text{накопленная}} = k_1 * O_{\text{текущий}} + k_2 * O_{\text{ауд}} + k_3 * O_{\text{сам. работа}}$$

где $O_{\text{текущий}}$ рассчитывается как взвешенная сумма всех форм текущего контроля, предусмотренных в ОУП

$$O_{\text{текущий}} = n_1 * O_{\text{эссе}} + n_2 * O_{\text{к/р}} + n_3 * O_{\text{реф}} + n_4 * O_{\text{кол}} + n_5 * O_{\text{дз}} ;$$

[Оставьте те формы текущего контроля, которые предусмотрены в ОУП. сумма удельных весов должна быть равна единице: $\sum n_i = 1$] Способ округления накопленной оценки текущего контроля: [указывается способ – арифметический, в пользу студента, другое].

Результирующая оценка за дисциплину рассчитывается следующим образом:

1. Если дисциплина преподается один модуль:

$$O_{\text{результ}} = k_1 * O_{\text{накопл}} + k_2 * O_{\text{экз}}$$

Способ округления накопленной оценки промежуточного (завершающего) контроля: [указывается способ – арифметический, в пользу студента, другое].

2. Если дисциплина преподается несколько модулей (например, 3):

$$O_{\text{промежуточная } i} = m_1 * O_{\text{текущая } i \text{ этапа}} + m_2 * O_{\text{промежуточный экзамен}}$$

Где $O_{\text{текущая } i \text{ этапа}}$ рассчитывается по приведенной выше формуле

$$O_{\text{накопленная завершающая}} = (O_{\text{промежуточная } 1} + O_{\text{промежуточная } 2} + O_{\text{накопленная } 3}) : \text{на число модулей}$$



Где $O_{\text{промежуточная 1}} + O_{\text{промежуточная 2}}$ – промежуточные оценки этапов 1 и 2,
а $O_{\text{накопленная 3}}$ – накопленная оценка последнего этапа перед завершающим экзаменом

Способ округления накопленной оценки промежуточного (завершающего) контроля в форме экзамена: [указывается способ – арифметический, в пользу студента, другое].

[Сумма удельных весов должна быть равна единице: $\sum m_i = 1$, при этом, $0,2 \leq m_i \leq 0,8$ (согласно Положению об организации промежуточной аттестации и текущего контроля успеваемости студентов НИУ ВШЭ, утвержденному приказом НИУ ВШЭ от 19.08.2014 №6.18.1-01/1908-02)

ОПЦИОНАЛЬНО: На экзамене студент может получить дополнительный вопрос (дополнительную практическую задачу, решить к передаче домашнее задание), ответ на который оценивается в 1 балл.

[Оставьте те оценки, которые учитываются при выставлении результирующей оценки за промежуточный или завершающий контроль. Сумма удельных весов должна быть равна единице: $\sum k_i = 1$, при этом, $0,2 \leq k_i \leq 0,8$ **После всех формул в обязательном порядке приводится способ округления полученного результата.**]

[Только для многомодульных дисциплин, по которым предусмотрен промежуточный контроль, укажите один из предложенных вариантов формирования оценки, которая идет в диплом]