

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН

НАО «Костанайский
региональный
университет
имени А.Байтурсынова»



Утверждаю
И.о. Проректора
А.Исмаилов
18.09.2020г.

ПРОГРАММЫ КУРСОВ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

ТЕХНОЛОГИИ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТА И ПЕРСПЕКТИВЫ РАЗВИТИЯ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

Костанай, 2020



Co-funded by the
Erasmus+ Programme
of the European Union

Программа повышения квалификации и переподготовки кадров лингвистического направления

ТЕХНОЛОГИИ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТА И ПЕРСПЕКТИВЫ РАЗВИТИЯ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

Общий объем – 36 часов. Сроки обучения – 2 недели.

Руководитель программы: Сатмаганбетова Ж. З.; м.т.н., ст. преподаватель кафедры программного обеспечения

Преподаватели: д.ф-м.н., доцент Медетов Н.А., к.ф-м.н., профессор Майер Ф.Ф., к.т.н., доцент Салыкова О.С., к.п.н. доцент Кандалина Е.М., к.ф.н. доцент Жабаева С.С., ст. преподаватель Бижанова О. И., ст. преподаватель Кулбаева М. М.

Цель курса: Обучить слушателей технологии автоматической обработки текста.

Задачи:

1. Систематизировать понятия и терминологию, используемые в области прикладного языкознания и компьютерной лингвистики;
2. Научить применять современные подходы, методы, инструменты, технологии в области компьютерной лингвистики, связанные с моделированием элементов искусственных и естественных языков;
3. Научить слушателей пользоваться с программными инструментальными средствами для решения различных задач, связанных с обработкой текстов.

Материально-техническая база: компьютерный класс с мультимедийным проектором и интернет.

Ожидаемый результат: слушатели освоили и систематизировали последние тенденции развития компьютерной лингвистики, выработали навыки применения компьютерных технологий обработки текстов на естественном языке, умеют решать лингвистические задачи, связанные с моделированием элементов искусственных и естественных языков.

Темы занятий:	Компьютерная лингвистика(КЛ, Computational Linguistics). Компьютерная лингвистика и искусственный интеллект. Разделы компьютерной лингвистики. Связь компьютерной лингвистики с другими областями знаний Искусственные языки. Классификация искусственных языков. Языки программирования. Формальные методы описания искусственных языков. Формальная грамматика. Понятие метаязыка. Информационный поиск(Information Retrieval) Характеристики
----------------------	---

	информационного поиска. Языки запросов. Языки индексирования. Вопросно-ответные системы. Поиск близких текстов (документов). Кластеризация и классификация текстов. Информационно-поисковые система(ИПС). Компьютерная лексикография (Computational Lexicography). Современные тенденции развития компьютерной лексикографии. Машинный перевод(МП). Виды МП. Технологии МП. Перспективы развития систем машинного перевода Извлечение информации из текстов. Автоматическое резюмирование. Контентный анализ. Синтез текстов. Распознавание синтез речи. Компьютерные программы для разметки и лингвистического анализа звука Корпусная лингвистика. Требования к корпусу текстов. Виды корпусов, принципы организации, системы управления, использование в прикладных целях. Разработка корпусов текстов. Этапы лингвистического анализа: предварительная обработка текста (Basic text processing); морфологический анализ(Part-Of-Speech Tagging, POS Tagging); Синтаксический анализ (Parsing, Syntactic parsing); Семантический анализ (Semantic parsing). Программные инструменты для лингвистического анализа и обработки текстов. Анализатор АОТ. Моделирование в компьютерной лингвистике. Лингвистические ресурсы. Платформы для разработки лингвистических процессоров. Проектирование лингвистических ресурсов и систем
--	---