

**Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
«Национальный исследовательский университет
„Высшая школа экономики“»**

Научно-учебная лаборатория лингвистической конфликтологии
и современных коммуникативных практик
Школы филологии

**Майнор «Язык и новые коммуникативные практики»
Рабочая программа дисциплины «Компьютерная лингвистика»**

для уровня подготовки — бакалавриат

Разработчик(и) программы

Пиперски Александр Чедович, к. ф. н., н. с., apiperski@hse.ru

Сомин Антон Александрович, м. н. с., somin@tut.by

Утверждена «08» февраля 2016 г.

Руководитель Научно-учебной лаборатории лингвистической конфликтологии и
современных коммуникативных практик Школы филологии

М. А. Кронгауз _____

Москва, 2015

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

Читается: 1–2 модуль, 3-й курс

Пререквизиты: нет

Трудоемкость: 76 часов (38 часов лекций, 38 часов семинаров)

Формы контроля: экзамен (устный), 1 контрольная работа, 1 домашнее задание

О курсе

Компьютерные технологии и наука о языке связаны очень тесно. С одной стороны, лингвисты используют компьютерные технологии в своих исследованиях; с другой стороны, лингвистические модели необходимы для автоматической обработки естественного языка, без которой наше взаимодействие с компьютером было бы совсем другим. Это и определяет структуру курса. В первой его части обсуждаются те компьютерные инструменты, которые могут использоваться в лингвистических исследованиях, в первую очередь корпуса разных типов, а во второй части речь идёт про задачи и технологии автоматической обработки текста: машинный перевод, автоматическое извлечение информации и т. д. Предлагается необходимый репертуар программных и математических ресурсов, однако специальная математическая подготовка не требуется — достаточно знаний в рамках школьной программы (а то, что вы из нее забыли, мы специально повторим). Кроме того, в результате курса студенты получают представление о применении статистических методов в гуманитарных науках.

Примерный перечень тем

1. Представление языковых данных в компьютерном виде.
2. Лингвистические корпуса и их применение в лингвистических исследованиях.
3. Статистическое исследование языковых явлений.
4. Основные задачи автоматической обработки естественного языка.
5. Машинный перевод и его оценка.
6. Исправление опечаток.
7. Автоматический грамматический разбор текстов на естественном языке.
8. Автоматическое извлечение информации из текстов.
9. Классификация и кластеризация текстов.
10. Распознавание письменной и устной речи.

Базовые учебники

Копотев, М. В. 2014. *Введение в корпусную лингвистику*. Praha: Animedia Company.
Manning, Christopher D. & Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. Cambridge, Mass: MIT Press.