

Учреждение образования  
«Брестский государственный университет имени А.С. Пушкина»

**УТВЕРЖДАЮ**

Проректор по учебной работе

Учреждения образования

«Брестский государственный

университет имени А.С. Пушкина»

\_\_\_\_\_ Е.Д. Осипов

«30» июня 2016 г.

Регистрационный № УД-26-007-16/уч.

**ОСНОВЫ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ**

**Учебная программа учреждения высшего образования  
по учебной дисциплине для специальности:**

1-21 06 01-01

Современные иностранные (немецкий, английский)  
языки (преподавание)

2016 г.

Учебная программа составлена на основе образовательного стандарта высшего образования первой ступени специальности 1-21 06 01-2013 «Современные иностранные языки (по направлениям)», утвержденного и введенного в действие постановлением Министерства образования Республики Беларусь от 30.08.2013г., № 88.

**СОСТАВИТЕЛЬ:**

**М.П. Концевой**, старший преподаватель кафедры прикладной математики и информатики физико-математического факультета Учреждения образования «Брестский государственный университет имени А.С. Пушкина»

**РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:**

Кафедрой прикладной математики и информатики Учреждения образования «Брестский государственный университет имени А.С. Пушкина»

(протокол № \_\_\_ от \_\_\_\_\_ 2016 г.)

Учебно-методической комиссией физико-математического факультета Учреждения образования «Брестский государственный университет имени А.С. Пушкина»

(протокол № \_\_\_ от \_\_\_\_\_ 2016 г.)

Согласовано

Декан факультета иностранных языков

С.Н. Северин

Научно-методическим советом Учреждения образования «Брестский государственный университет имени А.С. Пушкина»

(протокол № \_\_\_ от \_\_\_\_\_ 2016 г.)

## ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Компьютерная лингвистика является важнейшим структурным компонентом искусственного интеллекта, современных информационно-коммуникационных технологий, методов автоматической обработки текста. Изучение проблематики компьютерной лингвистики позволяет актуализировать на новом материале методы и результаты данных научно-технических областей и открывает новые подходы к изучению языка.

**Цель преподавания:** формирование у студентов лингвистических и технологических компетенций в области компьютерной лингвистики.

### **Задачи:**

- овладение студентами системой знаний о научных основах компьютерного анализа и обработки текста, его предмете, задачах, методах и путях их решения, инструментарии и приложениях;
- приобретение студентами опыта в области постановки и решения профильных задач на основе использования современных языковых и компьютерных технологий, навыков работы с современными системами и сервисами автоматической обработки текста;
- освоение студентами основных методов и средств автоматизации обработки текстовой информации, получение навыков комплексного использования средств вычислительной техники в решении задач автоматизации обработки текста;
- приобретение начального опыта научного анализа информационной составляющей текстовых сообщений и поиска путей их решения на основе использования систем и средств компьютерной обработки.

Освоение учебной программы должно обеспечить формирование следующих групп компетенций:

### **академических:**

- уметь применять базовые научно-теоретические знания для решения теоретических и практических задач;
- владеть системным и сравнительным анализом;
- уметь работать самостоятельно;
- владеть междисциплинарным подходом при решении проблем;
- иметь навыки, связанные с использованием технических устройств, управлением информацией и работой с компьютером;
- обладать навыками устной и письменной коммуникации;
- уметь учиться, повышать свою квалификацию в течение всей жизни;
- уметь осуществлять учебно-исследовательскую деятельность;
- уметь регулировать образовательные отношения и взаимодействия в педагогическом процессе;

### **социально-личностных:**

- обладать качествами гражданственности;
- быть способным к социальному взаимодействию;

- обладать способностью к межличностным коммуникациям;
- владеть навыками здоровьесбережения;
- быть способным к критике и самокритике;
- уметь работать в команде;
- быть способным к осуществлению самообразования и самосовершенствования профессиональной деятельности;
- быть способным к реализации ценностно-ориентационной деятельности;

**профессиональных:**

- эффективно реализовывать языковую деятельность;
- использовать оптимальные методы, формы, средства работы с текстом;
- осуществлять самообразование и самосовершенствование профессиональной деятельности с учетом современных образовательных, лингвистических технологий и технологических инноваций.

В результате изучения дисциплины студенты **должны знать:**

- научные и общекультурные основы возникновения компьютерной лингвистики;
- принципы и возможности моделирования языка и текста в коммуникативных системах;
- способы представления знаний о связном тексте в современных информационных интеллектуальных системах;
- типологии и лингвистические принципы создания систем и сервисов автоматической обработки текста;
- тенденции развития и основные направления современных методов автоматической обработки текста и систем искусственного интеллекта;
- понятие, задачи, технологии и правила использования языковых корпусов и баз данных;
- средства и сервисы поддержки приложений компьютерного анализа и обработки текста в современной информационной среде;

**должны уметь:**

- эффективно использовать аппаратное и программное обеспечение генерации, анализа, обработки и воспроизведения текстовой информации;
- анализировать и оценивать алгоритмы обработки текстовой информации;
- эффективно использовать программное обеспечение автоматической обработки связного текста;
- осуществлять лингвистическую, семантическую и гипертекстовую разметку текстовой информации;
- создавать текстовые материалы на основе иных семиотических систем;
- извлекать и обрабатывать метаязыковую информацию из текстовых сообщений.

На практических занятиях в курсе «Основы компьютерной лингвистики» рассматриваются теоретические вопросы и решаются конкретные задачи из различных областей теоретической и прикладной лингвистики, компьютерных технологий, осуществляется контроль знаний студентов. В самостоятельную работу студентов входит изучение и проработка теоретического материала, решение задач, написание рефератов, подготовка к практическим занятиям.

**Используемые формы и методы обучения и воспитания:** практические работы, защита рефератов. Преподавание курса осуществляется в локальной компьютерной сети (ЛКС) с использованием ЭОС (электронной образовательной среды), обеспечивающей возможность самостоятельной управляемой работы студентов и предполагающей применение форм и методов индивидуального и группового компьютерного обучения (интерактивного, мультимедийного, проблемного).

#### **Связь дисциплины с другими дисциплинами**

Для изучения дисциплины «Основы компьютерной лингвистики» студент должен предварительно владеть базовыми понятиями и навыками из курсов «Основы информационных технологий», «Основы семиотики», «История письменности», «Методы автоматической обработки текста».

Изучение дисциплины «Основы компьютерной лингвистики» осуществляется в тесной связи с тематикой следующих учебных дисциплин: «Общее языкознание», «Основы перевода», «Теория порождения речи», «Компьютерный перевод», «Компьютерное моделирование».

Форма получения образования дневная.

Семестр (курс)	Общее количество часов	Аудиторное количество часов	Лекции	Лабораторные занятия	УСР
Семестр 7 (4 курс)	130	66	16	38	12
итого	130	66	16	38	12

В соответствии со стандартом специальности программа дисциплины «Основы компьютерной лингвистики» предусматривает для изучения 130 часов, из которых – аудиторных часов 66 (16 часов лекционных, 38 часов практических, 12 часов УСР). Основной отчетной формой изучения дисциплины является зачет.

## **СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА**

### **Тема 1 Понятие компьютерной лингвистики**

Термин «компьютерная лингвистика» в различных терминологических системах. Подходы к определению компьютерной лингвистики через объект исследования, предмет, методы, инструментарий. Компьютерная лингвистика как наука и область исследовательских и конструкторских разработок. Компьютерная лингвистика в ряду других лингвистических дисциплин, ее цели и задачи. Предмет и объект компьютерной лингвистики.

Компьютерная лингвистика как учебный предмет для филологических и математических специальностей. Основные центры изучения компьютерной лингвистики. Образовательные ресурсы по компьютерной лингвистике.

### **Тема 2 Научные основы компьютерной лингвистики**

Эволюция научного понимания языка и задач его изучения (В. Фон Гумбольдт, Август Шлейхер, Бодуэн де Куртенэ, Ф. де Соссюр). Научные метафоры языка. Лингвистический поворот. Л. Витгенштейн и структурная лингвистика. Генеративные грамматики Н. Хомского, теория «Смысл-Текст» И. А. Мельчука. Формальная семантика Ричарда Монтегю.

Современная математика в периодизации А. Колмогорова. Логические и лингвистические проблемы современной математики. Языки формальной логики. Математическая логика. Формальные грамматики Исчисления. Предикаты. Нечеткая логика. Лингвистические вычисления. Проблема вычислимости и сложности. Математическая лингвистика. Компьютер как языковая машина. Лингвистические проблемы компьютерных наук. Идеи А. Тьюринга, А. Чёрча, А.Н. Колмогорова, А.А. Маркова. Математические выражения грамматик Хомского. Формальные языки. Языки программирования и разметки.

### **Тема 3 Квантитативная лингвистика**

Понятие квантитативных методов. Основные статистические категории.

Статистико-комбинаторные, дистрибутивно-статистические и дешифровочные методы в грамматике. Меры автономности грамматических структур и их семантики. Функциональная нагрузка грамматических категорий.

Частотные словари, словари-индексы, машинные фонды лексики. Конкордансы. Регулярные выражения. Характеристики частотных словарей и индексы покрытия текста. Квантитативные закономерности. Ограниченность квантитативных методов исследования языка.

### **Тема 4 Приложения компьютерной лингвистики**

Графематический и фонетический анализ. Морфологический анализ. Синтаксический анализ. Семантический анализ и онтологии.

Автогенерация текста. Атрибуция текста. Фоносемантический анализ, Semantic Web, системы и сервисы. Распознавание и синтез устной речи. Компьютерная лексикография и перевод. Методы поддержания диалога на естественных языках. Автоматическая обработка текста в системах искусственного интеллекта.

### **Тема 5 Корпусная лингвистика**

Понятие лингвистического корпуса. История корпусной лингвистики. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой.

Понятие разметки. Метаразметка. Языки разметки. Метаязык. Лингвистическая разметка (аннотирование). Задачи и основные направления корпусной лингвистики. Развитие лингвистических корпусов в мире.

Типы корпусов. Лингвистическая аннотация и метаданные. Лингвистические исследования на базе корпуса.

Методы извлечения информации из корпуса. Типы извлекаемой информации. Конкорданс.

Программы для работы с корпусом.

## УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Средства обучения (оборудование, учебно-наглядные пособия и др.)	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Количество часов УСР		
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
	<b>Основы компьютерной лингвистики (66 ч.)</b>	<b>16</b>	<b>38</b>			<b>12</b>	ПК, ВС, ЭОС	<b>зачёт</b>
<b>1</b>	<b>Понятие компьютерной лингвистики (2 ч.)</b>	<b>2</b>					ПК, ВС, ЭОС	Тест
<b>2</b>	<b>Научные основы компьютерной лингвистики (4 ч.)</b>	<b>4</b>	<b>4</b>				ПК, ВС, ЭОС	Тест, защита реферата
<b>3</b>	<b>Квантитативная лингвистика (14 ч.)</b>	<b>2</b>	<b>10</b>			<b>2</b>		



3.1	Понятие количественных методов	2					ПК, ВС, ЭОС	Тест
3.2	Статистические методы в грамматике		2				ПК, ВС, ЭОС	Отчет по пр.раб.
3.3	Частотные словари		4			2	ПК, ВС, ЭОС	Отчет по пр.раб., защита реферата
3.4	Регулярные выражения		4				ПК, ВС, ЭОС	Отчет по пр.раб.
<b>4</b>	<b>Приложения компьютерной лингвистики (26 ч.)</b>	<b>4</b>	<b>12</b>			<b>10</b>		
4.1	Компьютерный анализ текста		2			2	ПК, ВС, ЭОС	Отчет по пр.раб.
4.2	Семантический анализ и онтологии		2				ПК, ВС, ЭОС	Отчет по пр.раб.
4.3	Распознавание и синтез устной речи		2			2	ПК, ВС, ЭОС	Отчет по пр.раб.
4.4	Компьютерная лексикография и перевод	2	2			2	ПК, ВС, ЭОС	Отчет по пр.раб.
4.5	Методы поддержания диалога на естественных языках		2			2	ПК, ВС, ЭОС	Отчет по пр.раб.
4.6	Автоматическая обработка текста в системах искусственного интеллекта	2	2			2	ПК, ВС, ЭОС	Отчет по пр.раб., защита реферата,

<b>5</b>	<b>Корпусная лингвистика (16 ч.)</b>	<b>4</b>	<b>12</b>					
5.1	Понятие лингвистического корпуса	2					ПК, ВС, ЭОС	Тест, защита реферата
5.2	Метаязыковая разметка, лингвистическое аннотирование	2	4				ПК, ВС, ЭОС	Отчет по пр.раб.
5.3	Лингвистические исследования на основе языковых корпусов		4				ПК, ВС, ЭОС	Отчет по пр.раб.
5.4	Программы для работы с корпусами		4				ПК, ВС, ЭОС	Отчет по пр.раб.
	<b>ВСЕГО ЧАСОВ:</b>	<b>16</b>	<b>38</b>			<b>12</b>		<b>зачёт</b>

## ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

### ЛИТЕРАТУРА

#### Основная:

1. Зубов, А.В., Зубова, И.И. Основы искусственного интеллекта для лингвистов: учебное пособие. – М.: Университетская книга; Логос, 2007. – 320 с.
2. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / Большакова Е.И. [и др.]. – М.: МИЭМ, 2011. – 272 с.
3. Боярский, К.К. Введение в компьютерную лингвистику: учебное пособие. – СПб: НИУ ИТМО, 2013. – 72 с.
4. Компьютерная лингвистика: научное направление и учебная дисциплина: сборник научных статей. Вып. 1 / В. И. Коваль (ответств. ред.) [и др.]; М-во образования РБ, ГГУ им. Ф. Скорины. – Гомель: ГГУ им. Ф. Скорины, 2010. – 236 с.

#### Дополнительная:

5. Пентус, А.Е., Пентус М.Р. Теория формальных языков: учебное пособие. – М.: МГУ, 2004. – 80 с.
6. Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 26–30 мая 2010 г.). Вып. 9 (16). – М.: Изд-во РГГУ, 2010.
7. Лингвистика для всех. Летние лингвистические школы 2005 и 2006 / Ред.–сост. Е.В. Муравенко, О.Ю. Шеманаева. – М.: МЦНМО, 2008. – 440 с.
8. Зализняк А.А. Из заметок о любительской лингвистике. – М.: Русский Мирь, 2008. – 240 с.
9. Волкова, И.А. Введение в компьютерную лингвистику. Практические аспекты создания лингвистических процессоров: учебное пособие. – М., 2006. – 43 с.
10. Апресян, Ю.Д. Идеи и методы современной структурной лингвистики. – М.: Просвещение, 1966 г. – 34 с.
11. Апресян, Ю.Д. Лингвистический процессор для сложных информационных систем / Ю.Д. Апресян [и др.]. – М.: Наука, 1992. – 256 с.
12. Мельчук, И.А. Опыт теории лингвистических моделей «Смысл Текст». – М.: Школа «Языки русской культуры», 1999. – 346 с.
13. Баранов, А.Н. Введение в прикладную лингвистику: учебное пособие. – М.: Эдиториал УРСС, 2001. – 360 с.
14. Шемякин, Ю.И. Начала компьютерной лингвистики: учеб. пособие. – М.: Изд-во МГОУ, А/О «Росвузнаука», 1992. – 53 с.
15. Анисимов, А. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык. – Киев: Наук. думка, 1991.
16. Лейчик, В.М. Терминоведение: предмет, метод, структура. – М., 2007, – 256 с.

- 17.Гринев-Гриневиц, П.В. Введение в терминографию: как просто и легко составить словарь : учебное пособие. – М.: Книжный дом ЛИБРОКОМ, 2009, – 224 с.
- 18.Келлер, Р. Языковые изменения. О невидимой руке в языке / Пер. с нем. и вступ. ст. О.А. Костровой. – Самара: Издательство СамГПУ, 1997. – 312 с.
- 19.Белоногов, Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004. – 248 с.

### **Примерные перечни заданий управляемой самостоятельной работой**

**Тема:** Частотные словари (2 ч.)

**Форма контроля:** защита реферата.

**Задание:** подготовить реферат в виде защищенного паролем PDF-документа, содержащего структурированное сравнительное описание статистики предложенных преподавателем текстов и их частотных словарей, аргументированные заключения о следующих параметрах данных текстов: богатство речи, Fog Index, Flesch Reading Ease.

**Средства:** текстовые процессоры, ПО для статистической обработки текста

**Источники информации:** <http://lab314.brsu.by/kmp-lite/kmp2.htm>

**Тема:** Компьютерный анализ текста (2 ч.)

**Форма контроля:** защита реферата.

**Задание:** подготовить реферат в виде защищенного паролем PDF-документа, содержащего структурированное сравнительное описание двух поисковых систем (предложенных преподавателем), а также языков поисковых запросов используемых в них, оценку результатов работы данных систем в отношении полноты поиска, релевантности и пертинентности для двух языков (предложенных преподавателем).

**Средства:** поисковые системы в Интернете

**Источники информации:** <http://lab314.brsu.by/kmp-lite/kmp2.htm>

**Тема:** Компьютерная лексикография и перевод (2 ч.)

**Форма контроля:** защита реферата.

**Задание:** подготовить реферат в виде защищенного паролем PDF-документа, содержащего оригинальный текст (предложенных преподавателем), а также версию текста подготовленного к переводу, результат машинного перевода текста (на русский или белорусский язык), результат постредактирования машинного перевода, оценку результатов работы компьютерного переводчика (предложенного преподавателем) в отношении различных уровней эквивалентности текста).

**Средства:** онлайн-системы и сервисы перевода, Promt, Belazar

**Источники информации:** <http://lab314.brsu.by/kmp-lite/kmp2.htm>

**Тема:** Распознавание и синтез устной речи (2 ч.)

**Форма контроля:** защита реферата.

**Задание:** подготовить реферат в виде защищенного паролем PDF-документа, содержащего структурированное сравнительное описание двух онлайн-сервисов распознавания устной речи (предложенных преподавателем), а также оценку результатов работы данных сервисов в отношении передачи лексических, семантических, синтаксических и экстралингвистических параметров текста.

**Средства:** сервисы распознавания речи в Интернете

**Источники информации:** <http://lab314.brsu.by/kmp-lite/kmp2.htm>

**Тема:** Методы поддержания диалога на естественных языках (2 ч.)

**Форма контроля:** защита реферата.

**Задание:** подготовить реферат в виде защищенного паролем PDF-документа, содержащего структурированное сравнительное описание двух онлайн-диалоговых ботов (предложенных преподавателем), а также языков разметки диалога используемых в них, оценку результатов работы данных систем в отношении правдоподобности диалога на две (предложенные преподавателем) темы.

**Средства:** диалоговые системы в Интернете

**Источники информации:** <http://lab314.brsu.by/kmp-lite/kmp2.htm>

**Тема:** Автоматическая обработка текста в системах искусственного интеллекта (2 ч.)

**Форма контроля:** защита реферата.

**Задание:** подготовить реферат в виде защищенного паролем PDF-документа, содержащего структурированное описание системы искусственного интеллекта (предложенной преподавателем), а также ее языкового обеспечения, оценку результатов работы данной системы.

**Средства:** интеллектуальные системы в Интернете

**Источники информации:** <http://lab314.brsu.by/kmp-lite/kmp2.htm>

**Перечень используемых средств диагностики результатов учебной деятельности:**

1. Тесты;
2. Отчеты по практическим работам;
3. Защита рефератов;
4. Контрольный опрос;
5. Зачет.

## **Средства обучения (оборудование, учебно-наглядные пособия и др.)**

Персональный компьютер, локальная вычислительная сеть УО «Брестский гос. ун-т имени А.С. Пушкина», электронная образовательная среда «ЭОС», интернет-ресурсы:

1. Общие курсы. Русская морфология :: МГУ :: Филфак :: ОТиПЛ  
<http://www.philol.msu.ru/~otipl/new/main/courses/modelling.php>
2. AINL: Искусственный Интеллект и Естественный Язык  
<http://nlpseminar.ru/ainl/>
3. Программа поиска наиболее устойчивых словосочетаний (коллокаций) - новая версия [alingva.ru/index.php/lingvosoft/18-key-word](http://alingva.ru/index.php/lingvosoft/18-key-word)
4. Диалог: сайт конференции по компьютерной лингвистике  
[www.dialog-21.ru](http://www.dialog-21.ru)
5. OpenCorpora: открытый корпус русского языка [opencorpora.org](http://opencorpora.org)
6. Национальный корпус русского языка <http://corpora.yandex.ru/>,  
<http://www.ruscorpora.ru/>
7. Corpus Albaruthenicum <http://grid.bntu.by/corpus/>
8. LingLing.ru — Лингвистика для школьников [www.lingling.ru/Introduction/](http://www.lingling.ru/Introduction/)
9. Интерлингвистика и искусственные языки на сайте Игоря Гаршина. Всеобщий язык [garshin.ru/linguistics/model/index.html](http://garshin.ru/linguistics/model/index.html)
10. Natural Language Processing | Программа и материалы конференции [nlpseminar.ru/ainl/program/](http://nlpseminar.ru/ainl/program/)
11. Программа склонения по падежам [morpher.ru](http://morpher.ru)
12. Link Grammar for Russian - Грамматика связей для русского языка - Разбор предложения <http://www.sz.ru/parser/>
13. Система лингвистического поиска для русского, английского и немецкого языков, доступная с сайта <http://aot.ru/>
14. Филология и лингвистика (компьютерная лингвистика)  
<http://www.filologia.su/kompjuternaja/>
15. Компьютерная лингвистика на сайте Игоря Гаршина <http://www.garshin.ru>
16. Портал знаний по компьютерной лингвистике <http://www.niserv.iis.nsk.su>
17. Актуальные проблемы компьютерной лингвистики на сайте Научные семинары АБВУУ <http://www.abbyu.ru/science/seminars/archive/>
18. Онлайн-овая лингвистическая лаборатория Microsoft  
<http://labs.microsofttranslator.com/>
19. Проект «Сеть словесных ассоциаций» <http://wordassociations.ru/>

## **ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ**

MS Windows, MS Internet Explorer, Google Chrome, MS Word, MS Excel, 7Zip, Prompt, АБВУУ Lingvo, Belazar, Unique Record Set Management, Balabolka, АБВУУ FineReader, Vaal mini, MS Front Page, Acrobat Reader, FDS, XML Notepad++, MAXQUDA, . Syntext Serna, Fluent Editor, RitmMe, XML Notepad, FictionBook Editor, Calibre.



## ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ УВО

на \_\_\_\_/\_\_\_\_ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры прикладной математики и технологий программирования Учреждения образования «Брестский государственный университет имени А.С. Пушкина»

(протокол № \_\_\_\_ от \_\_\_\_\_ 20\_\_ г.)

Заведующий кафедрой

\_\_\_\_\_

(ученая степень, ученое звание)

\_\_\_\_\_

(подпись)

\_\_\_\_\_

(И.О.Фамилия)

УТВЕРЖДАЮ

Декан факультета

\_\_\_\_\_

(ученая степень, ученое звание)

\_\_\_\_\_

(подпись)

\_\_\_\_\_

(И.О.Фамилия)