**Министерство образования Республики Беларусь**

**Учебно-методическое объединение по гуманитарному образованию**

**УТВЕРЖДАЮ**

Первый заместитель Министра образования

Республики Беларусь

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А. И. Жук

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2012 г.

Регистрационный № ТД-\_\_\_\_\_\_\_\_/ тип.

**МЕТОДЫ АВТОМАТИЧЕСКОЙ**

**ОБРАБОТКИ ТЕКСТА**

Типовая учебная программа

для высших учебных заведений по специальностям

1-21 05 01 «Белорусская филология» (по направлениям)

Направление специальности 1-21 05 01-02 «Белорусская филология (компьютерное обеспечение)»

1-21 05 02 «Русская филология» (по направлениям)

Направление специальности 1-21 05 02-02 «Русская филология   
(компьютерное обеспечение)»

СОГЛАСОВАНО СОГЛАСОВАНО

Председатель Начальник Управления

Учебно-методического объединения высшего и среднего специального образования

по гуманитарному образованию Министерства образования Республики Беларусь

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В. Л. Клюня \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ С. И. Романюк

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2012 г. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2012 г.

Проректор по научно-методической работе

Государственного учреждения образования

«Республиканский институт высшей школы»

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ И. В. Титович

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2012 г.

Эксперт-нормоконтролер

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2012 г.

Минск 2012

**СОСТАВИТЕЛИ:**

**Рычкова Людмила Васильевна** – заведующий кафедрой общего и славянского языкознания учреждения образования «Гродненский государственный университет имени Янки Купалы», кандидат филологических наук, доцент

**Станкевич Алеся Юрьевна** – старший преподаватель кафедры общего и славянского языкознания учреждения образования «Гродненский государственный университет имени Янки Купалы»

**РЕЦЕНЗЕНТЫ:**

кафедра интеллектуальных информационных технологий факультета информационных технологий и управления учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»;

**Карпилович Татьяна Павловна** – заведующий кафедрой теории и практики английского языка переводческого факультета учреждения образования «Минский государственный лингвистический университет», доктор филологических наук, профессор

**РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ В КАЧЕСТВЕ ТИПОВОЙ:**

кафедрой общего и славянского языкознания филологического факультета учреждения образования «Гродненский государственный университет имени Янки Купалы» (протокол № 5 от 11 мая 2012 г.);

Научно-методическим советом Белорусского государственного университета (протокол № 5 от 28 мая 2012 г.);

Научно-методическим советом по филологическим специальностям учебно-методического объединения по гуманитарному образованию (протокол № 6 от 21 июня 2012 г.);

**Ответственный за редакцию**: Л. В. Рычкова

**Ответственный за выпуск**: Л. В. Рычкова

**1. ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

**1.1. Характеристика учебной дисциплины**

Дисциплина «Методы автоматической обработки текста» входит в состав обязательных дисциплин, предусмотренных образовательными стандартами филологических специальностей в рамках направления «Компьютерное обеспечение». Содержательное ядро дисциплины составляет проблематика такой интердисциплинарной отрасли прикладных лингвистических знаний, как АОТ (автоматическая обработка текста). Предполагается, что студенты к моменту изучения данной дисциплины уверенно овладели навыками пользования типовым программным и информационным обеспечением, полученными ими при изучении дисциплин «Основы информатики» и «Введение в компьютерную филологию», а также овладели основным понятийным аппаратом как компьютерной лингвистики, так и «зонтичной» области компьютерной филологии. Овладение дисциплиной «Методы автоматической обработки текста» необходимо для непосредственной поддержки профессиональной деятельности специалиста по компьютерной филологии.

**1.2. Цель и задачи учебной дисциплины**

**Цель дисциплины**  – сформировать у студентов системное представление о классах методов АОТ, языковых и инструментальных средствах АОТ; выработать у них представления о направлениях применения методов и инструментария АОТ для решения профессиональных задач.

Основные **задачи** дисциплины включают:

– выработку представлений о классах методов автоматической обработки текстов, видах и перспективах АОТ (теоретическая составляющая);

– выработку практических навыков решения задач по разделам курса (собственно прикладная, или практико-ориентированная составляющая).

В результате изучения учебной дисциплины студент должен:

**знать**:

– терминологический аппарат изучаемой предметной области;

– сферы применения методов и средств АОТ, в том числе, для целей филологии;

– язык регулярных выражений как средства проверки, фильтрации текстовых данных и как средства управления текстовыми данными;

**уметь**:

– решать профессиональные задачи с обращением к методам и инструментам АОТ;

– работать с изучаемыми в пределах курса инструментами АОТ;

– уметь использовать регулярные выражения как средство проверки и фильтрации текстовых данных.

Основные **компетенции**, которые должны быть сформированы у студентов:

*– компетенция в сфере познавательной деятельности:* применение методов компьютерной лингвистики для решения профессионально-ориентированных задач;

*– профессиональные компетенции:* владение компьютерными методами сбора и обработки данных; владение приемами работы с программными средствами АОТ.

**1.3. Рекомендуемые формы и методы обучения и воспитания** включают:

- интерактивные лекции с использованием современных технических средств наглядности и разнообразных форм дидактического диалога;

- практические и лабораторные занятия, с включением подсистемы контролируемых самостоятельных заданий (в том числе, выполняемых в системах ДО).

Основной метод обучения: наглядно-практический.

**1.4. Рекомендации по организации самостоятельной работы студентов**

Формой проведения самостоятельной работы студентов, оптимальной при изучении данной дисциплины, является выполнение контрольно-обучающих заданий средствами заданного программного инструмента либо разработка проектов (компонентов систем АОТ).

Рекомендуемая форма итогового контроля знаний и умений студентов - экзамен.

**1.5. Рекомендуемый объем часов и формы проведения занятий**

На изучение дисциплины «Методы автоматической обработки текста» отводится 102 часа, из них аудиторных – 68 часов.

Примерное распределение аудиторного времени следующее:

– лекции – 10 часов;

– практические занятия – 10 часов;

– лабораторные занятия (в компьютерном классе) – 48 часов.

**2.** **ПРИМЕРНЫЙ ТЕМАТИЧЕСКИЙ ПЛАН**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Название раздела**  (темы, занятия) | **Количество аудиторных часов** | | | |
| **лекции** | **практические занятия** | **лабораторные занятия** | **семинары** | |
| **1** | **2** | **3** | **4** | **5** | **6** | |
| 1. | Классы методов АОТ | 1 |  |  |  | |
| 2.1 | Средства обработки текстовой информации | 1 |  | 12 |  | |
| 2.2 | Подъязык регулярных выражений | RegExp как средство обработки текстовой информации |  |  | 16 |  | |
| 3. | Автоматический морфологический анализ | МА | 2 | 2 | 4 |  | |
| 4. | Автоматический синтаксический анализ | СА | 2 | 2 | 4 |  | |
| 5.1. | Автоматическое индексирование |АИ | 1 | 2 | 4 |  | |
| 5.2. | Автоматическое извлечение терминов из текста | 1 | 2 | 4 |  | |
| 6. | Автоматическое реферирование|АР | 2 | 2 | 4 |  | |
|  | **ИТОГО ЧАСОВ ПО КУРСУ:** | **10** | **10** | **48** |  | |

**3. СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА**

**1. Классы методов АОТ.**

Виды автоматической обработки текста:автоматический морфологический анализ, автоматический синтаксический анализ, автоматическое индексирование, автоматическое реферирование, автоматическое рубрицирование, автоматический семантический анализ. Классы методов АОТ: статистические и вероятностные методы; лингвистически-ориентированные методы (методы с опорой на лингвистические базы данных, лингвистические базы знаний) комбинированные методы. Научные коллективы, разрабатывающие проблемы АОТ на постсоветском пространстве.

**2.1. Средства обработки текстовой информации.**

Инструментальные средства АОТ. Языки программирования высокого уровня, ориентированные на обработку текстовой информации (Perl, Ruby, Python, LISP). Обзор Python-библиотек для автоматической обработки текста.

**2.2.** **Подъязык регулярных выражений | RegExp как средство обработки текстовой информации**.

Регулярные выражения как запись представления о коллекции строк. Виды программного обеспечения, поддерживающего RegExp.

Введение в синтаксис регулярных выражений. Символьные классы; сокращенные записи символьных классов. Квантификаторы. Жадные и ленивые квантификаторы. Сохранения. Проверки. Флаги. Регулярные выражения как средство обработки текста.

On-line конструкторы регулярных выражений.

**3. Автоматический морфологический анализ | МА.**

Методы автоматического МА: декларативный, процедурный, комбинированный. Морфологические анализаторы. Методы разметки по частям речи (POS-разметки). Автоматическая лемматизация.

**4. Автоматический синтаксический анализ | СА.**

Методы СА. Модели синтаксической структуры для СА: графы зависимости, графы непосредственно составляющих. Синтаксические анализаторы.

**5.1. Автоматическое индексирование | АИ.**

Объекты АИ. Деривационное и приписное АИ. Методы АИ: статистические, словарные, морфологические, синтаксические, семантические, суперсинтаксические, позиционные. Метод концептуального индексирования (лингвистическая онтология для АИ).

**5.2. Автоматическое извлечение терминов из текста.**

Определение важности термина относительно коллекции документов: мера TF\*IDF, метод оценки различительной силы термина. Методы извлечения терминов, основанные на критериях связности.

**6. Автоматическое реферирование | АР.**

Автореферирование и квазиреферирование. Полуавтоматическое реферирование. Архитектура систем АР, работающих с опорой / без опоры на знания.

Методы автоматического реферирования: статистические, позиционные, суперсинтаксические, семантико-синтаксические. Классификация методов АР по Т.П. Карпилович: методы статистические, поверхностно-синтаксические, глубинно-синтаксические, концептуальные.

Метод **симметричного** реферирования (В.А. Яцко, В.С. Ступин).

Метод **маркерного** реферирования (Д.И. Блюменау, Н.И. Гендина, Д.Г. Лахути, В.П. Леонов). Аспектная сетка документа.

**4. ИНФОРМАЦИОННО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ**

**ПО ДИСЦИПЛИНЕ**

**4.1. Перечень рекомендуемой литературы**

**/ К РАЗДЕЛУ 1**

**/Основная**

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В.– М.: МИЭМ, 2011.– 272 с.
2. Евдокимова, И.С. Естественно-языковые системы.– Улан-Удэ, 2006.
3. Марчук, Ю.Н. Компьютерная лингвистика.– М., 2007.
4. Сокирко, А.В.Обзор зарубежных систем автоматической обработки текста, использующих поверхностно-семантическое представление, и машинных семантических словарей// НТИ.Сер.2. Информационные процессы и системы. –2000.–№ 12 [http://www.viniti.ru/cgi-bin/nti/nti.pl?action=show&year=2\_2000&issue=12&page=1].
5. Хорошевский, В.Ф. Извлечение информации из текстов на конференциях серии ДИАЛОГ: взгляд соседа по лестничной клетке // 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.– С. 538–548 [http://www.dialog-21.ru/dialog2010/materials/html/78.htm].

**/Дополнительная**

1. Автоматизация обработки текста // НТИ. Сер.2. Информационные процессы и системы.– 1981.– № 2.
2. Агеев, М.С. Автоматическая рубрикация текстов: методы и проблемы / М.С. Агеев, Б.В. Добров, Н.В. Лукашевич // Ученые записки Казанского Государственного Университета. Серия Физико-математические науки. – 2008. – Т. 150, Кн. 4. – C.25–40. [http://www.cir.ru/docs/ips/publications/2008\_kgu\_classif.pdf]
3. Андреев, Н.Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении.– Л., 1967.
4. Кузнецов, И.П Система выявления из документов значимой информации на основе лингвистических знаний в форме семантических сетей / И.П. Кузнецов, В.П. Кузнецов, А.Г. Мацкевич // Труды Международного семинара Диалог'2000 по компьютерной лингвистике и ее приложениям, Протвино, 1-5 июня 2000 года.– В 2-х тт.– Т. 2. Прикладные проблемы [http://www.dialog-21.ru/Archive/2000/Dialogue%202000-2/232.htm].
5. Методы автоматического анализа и синтеза текста. – Мн., 1985.
6. Пиотровский, Р.Г. Математическая лингвистика / Р.Г. Пиотровский, К.Б. Бектаев, А.А. Пиотровская.– М.: Высш. школа, 1977.
7. Рубашкин, В.Ш. Семантический (концептуальный) словарь для информационных технологий /В.Ш. Рубашкин, Д.Г. Лахути. Часть I // НТИ.Сер.2. Информационные процессы и системы. – 1998. –№1.– С. 19 – 24; Часть II // НТИ.Сер.2. Информационные процессы и системы.– 1999.–-№5.– С. 1–12; Часть III // НТИ.Сер.2. Информационные процессы и системы.– 2000.–№7.–С. 1– 9.
8. URL: http://www.aot.ru/onlinedemo.html (сайт группы АОТ, раздел программ АОТ on-line).
9. URL: http://nlpseminar.ru/archive/ (сайт семинара NLP (Санкт-Петербург), раздел презентаций и видеолекций).

**/ К РАЗДЕЛУ 2.1**

**/Основная**

1. Носков, А.А. Инструментальные системы разработки приложений по автоматической обработке текстов на естественном языке // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В.- М.: МИЭМ, 2011.-272 с. – С. 141–169.
2. Python Programming Language – Official Website.– http://www.python.org/.
3. Сузи, Р.А. Язык программирования Python.– http://www.intuit.ru/department/pl/python/.
4. NLPK: Natural Language Toolkit.– <http://www.nltk.org/>.

**/Дополнительная**

1. Лутц, М. Изучаем Python. – Спб.–М.:Символ-Плюс, 2009.
2. Саммерфилд, М. Программирование на Python 3.– Спб.–М.:Символ-Плюс, 2009.
3. Bird, S. Natural Language Processing with Python / S. Bird, E. Klein, E. Loper.- Tokio: O'Reilly,2009.
4. URL: http://alias-i.com/lingpipe/ (сайт LingPipe, программных библиотек для анализа ЕЯ).
5. URL: http://opennlp.apache.org/ (сайт OpenNLP, программных библиотек для анализа ЕЯ)
6. URL: http://opennlp.sourceforge.net/projects.html (каталог open-source инструментов обработки ЕЯ).
7. URL: http://www.aot.ru/download.php (сайт группы АОТ, страница программных библиотек и бинарных файлов).
8. URL: http://www.sil.org/linguistics/computing.html (каталог ПО и других ресурсов по прикладной лингвистике от SIL).

**/ К РАЗДЕЛУ 2.2**

**/Основная**

1. Саммерфилд, М. Регулярные выражения // Саммерфилд, М. Программирование на Python 3.– Спб.–М.:Символ-Плюс, 2009.
2. URL: http://gskinner.com/RegExr/, http://realcode.ru/regexptester/ (on-line конструкторы регулярных выражений).

**/Дополнительная**

1. Фридл, Дж. Регулярные выражения. – СПб.: Питер, 2003.
2. URL: http://www.linguis.ru/about (ресурс К. Селиванова, посвященный АОТ; реализация на Python).
3. Zacharski, R. Python for Linguists.– http://www.zacharski.org/books/python-for-linguists/.

**/ К РАЗДЕЛУ 3**

**/Основная**

1. Евдокимова, И.С. Естественно-языковые системы.– Улан-Удэ, 2006.– С. 65–74.
2. Клышинский, Э.С. Начальные этапы анализа текста // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В.- М.: МИЭМ, 2011.-272 с. – С. 106–140.
3. Марчук, Ю.С. Компьютерная лингвистика.– М., 2007.– С. 60–70.

**/Дополнительная**

1. Белоногов, Г.Г Языковые средства автоматизированных информационных систем | Г.Г. Белоногов, Б.А. Кузнецов.– М., 1983.
2. Белоногов, Г.Г. Автоматизация процессов накопления, поиска и обобщения информации /Г.Г. Белоногов, А.П. Новоселов. – М., 1979.
3. Волкова, И.А. Программный комплекс для лингвистической обработки текстов на русском языке / И.А. Волкова, М.О. Проскурня // Труды Международного семинара Диалог'2002 по компьютерной лингвистике и ее приложениям.- В 2-х тт.- Т. 2. Прикладные проблемы  
   [http://www.dialog-21.ru/materials/archive.asp?id=7530&y=2002&vol=6078].
4. Коваль, С.А. О сравнимости и эквивалентности компьютерных представлений морфологии // Компьютерная лингвистика и интеллектуальные технологии. Тр. междунар. конференции Диалог'2003 (Протвино, 11–16 июня 2003 г.) / Под ред. И. М. Кобозевой, Н. И. Лауфер, В. П. Селегея. М.: Наука, 2003.- С. 305–311 [http://www.dialog-21.ru/Archive/2003/Koval.pdf].
5. Ляшевская, О.Н. Оценка методов автоматического анализа текста: морфологические парсеры русского языка / О.Н. Ляшевская, И. Астафьева, А. Бонч-Осмоловская // 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.– С. 318–326 [http://www.dialog-21.ru/dialog2010/materials/html/49.htm].
6. Сокирко, А.В. Морфологические модули на сайте www.aot.ru.–http://www.aot.ru/docs/sokirko/Dialog2004.htm.
7. URL: http://axofiber.org.ru/projects/rm/; http://axofiber.no-ip.org/projects/rm/rmu.test.htm (Проект «Русская морфология»).
8. URL: http://company.yandex.ru/technology/products/mystem/mystem.xml (страница парсера Mystem).
9. URL: http://ru-eval.ru/ (форум «Оценка методов автоматического анализа текста: морфологические парсеры русского языка»).
10. URL: http://www.aot.ru/demo/morph.html (Проект Dialing Mogphology).
11. URL: http://www.aot.ru/download/LexGroup.rar2 (Русский морфологический словарь Дмитрия Григорьева).

**/ К РАЗДЕЛУ 4**

**/Основная**

1. Евдокимова, И.С. Естественно-языковые системы.– Улан-Удэ, 2006.– С. 75–88.
2. Компьютерный синтаксический анализ: описание моделей и направлений разработок / Г.Д. Карпова, Ю.К. Пирогова, Т.Ю. Кобзарева, Е.В. Макаэлян // Итоги науки и техники. Серия «Вычислительные науки». Том 6. – М., 1991. – 243 с.

**/Дополнительная**

1. Волкова, И.А. Введение в компьютерную лингвистику. Практические аспекты создания лингвистических процессоров. Уч. пособие для студ. ВМиК МГУ. – М.: 2006. – 43 с.
2. Гладкий, А.В. Синтаксические структуры естественного языка. ‑ М.: ЛКИ, 2007.
3. Иорданская, Л.Н. Автоматический синтаксический анализ. Т. 2. Межсегментный синтаксический анализ.– Новосибирск, 1967.
4. Мельчук, И.А. Автоматический синтаксический анализ. Том 1. Общие принципы. Внутрисегментный синтаксический анализ. – Новосибирск, 1964.
5. URL: <http://syntax.ru/synt.html> (страница on-line синтаксического парсера Е.Д. Ермолаева).
6. URL: <http://sz.ru/parser/> (страница on-line синтаксического парсера С. Протасова Russian Link Grammar Parser) .
7. URL: http://www.aot.ru/demo/synt.html (синтаксический анализатор от группы АОТ).

**/ К РАЗДЕЛУ 5.1**

**/Основная**

1. Добров, Б.В. Разработка лингвистической онтологии для автоматического индексирования текстов по естественным наукам / Б.В. Добров, Н.В. Лукашевич, М.Н. Синицын, В.Н. Шапкин // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды седьмой Всероссийской научной конференции (3-7 октября 2005 г., г. Ярославль). - Ярославль: ЯрГУ, 2005. – С. 70-79[http://www.cir.ru/docs/ips/publications/2005\_rcdl\_ontology.pdf].
2. Пащенко, Н.А. Проблемы автоматизации индексирования и реферирования / Н.А. Пащенко, Л.В. Кнорина, Т.В. Молчанова и др. // Итоги науки и техники. Сер. Информатика. – М.: ВИНИТИ, 1983.– Т.7.– С. 7–111.
3. Солтон, Дж. Динамические библиотечно-информационные системы. – М.: Мир, 1979.– С. 90–125, 462–484.
4. Ягунова, Е.В. Слово – коллокация – синтаксические конструкции – текст. Единица анализа и контекст // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В.- М.: МИЭМ, 2011.-272 с. – С. 141–169. – С. 17–43.

**/Дополнительная**

1. Бевзов, А.Н. Разработка методов автоматического индексирования текстов на естественном языке для информационно-поисковых систем //Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» ‑ RDLC'2008.-Дубна, 2008.– С.401-404.
2. Лукашевич, Н.В. Тезаурус для автоматического концептуального индексирования как особый вид лингвистического ресурса // Труды международного семинара «Диалог-2001».– Аксаково: 2001.– С. 273–279.
3. Лукашевич, Н.В. Тезаурус для автоматического рубрицирования и индексирования: разработка, структура, ведение / Н.В. Лукашевич, А.Д. Салий // НТИ. Сер. 2. – 1996. № 1. – С. 1-6.
4. Лукашевич, Н.В. Тезаурус русского языка для автоматической обработки больших текстовых коллекций / Н.В. Лукашевич, Б.В. Добров // Труды Международного семинара «Диалог'2002» / Под ред. А.С. Нариньяни.– М., 2002. ‑ Т.2. Прикладные проблемы. – С.338–346.  
   [http://www.dialog-21.ru/materials/archive.asp?id=7619&vol=6078&y=2002]

**/ К РАЗДЕЛУ 5.2**

**/Основная**

1. Браславский, П.П. Сравнение четырех методов автоматического извлечения двухсловных терминов из текста / П.П. Браславский, Е. Соколов // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006» / Под ред. Н.И. Лауфер, А.С. Нариньяни, В.П. Селегея.– М., 2006. – С. 88–94.
2. Солтон, Дж. Динамические библиотечно-информационные системы.– М.: Мир, 1979.–С. 90–102, 462–466.

**/Дополнительная**

1. Гельбух, А.Ф. Автоматический поиск и классификация однословных терминов в корпусе предметной области с использованием логарифмической меры сходства с неспециализированным корпусом / А.Ф. Гельбух, Г.О. Сидоров, Э. Лавин-Вийа, Л. Чанона-Эрнандес //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.– С. 82–89 [http://www.dialog-21.ru/dialog2010/materials/html/14.htm].
2. Захаров, В.П. Статистический метод выявления коллокаций / В.П. Захаров, М.В. Хохлова // Языковая инженерия: в поиске смыслов / Доклады семинара «Лингвистические информационные технологии в Интернете»: XI Всероссийская объединенная конференция «Интернет и современное общество». Под ред. В.Ш. Рубашкина и В.П. Захарова. – СПб., 2008 [http://www.infosoc.ru/2008/pdf\_CL/Zakharov&Hohlova.pdf].
3. Скороходько, Э.Ф. Информационная сложность термина в научном тексте (индикативные и информативные маркеры) // НТИ.Сер.2. Информационные процессы и системы.–1998.– № 7.– С. 13–22.
4. Скороходько, Э.Ф. Семантическая продуктивность и семантическая емкость слова в общеупотребительной и терминологической лексике: некоторые количественные характеристики // НТИ. Сер.2. Информационные процессы и системы.– 1997.– №2.– 1–13.
5. Скороходько, Э.Ф. Семантическая сложность слова (термина): сетевые параметры и коммуникативные характеристики // НТИ. Сер.2. Информационные процессы и системы. – 1995.– № 2.– С. 4–11.
6. Ягунова, Е.В. Природа коллокаций в русском языке. Опыт автоматического извлечения и классификации на материале новостных текстов / Е.В. Ягунова, Л.М. Пивоварова // НТИ. – Сер.2.– 2010.– № 6.– С. 30–40 [http://www.webground.su/services.php?param=priroda\_collac&part=priroda\_collac.htm].

**/ К РАЗДЕЛУ 6**

**/Основная**

1. Башмаков, А.И. Интеллектуальные информационные технологии / А.И. Башмаков, И.А. Башмаков.– М.: Изд-во МГТУ им. Н.Э. Баумана, 2005.– С. 77–90.
2. Блюменау, Д. И. Развитие индикаторного метода компьютерного свертывания текстов / Д.И. Блюменау, Л.Н. Афанасова // НТИ.Сер.2. Информационные процессы и системы.– 2002. – № 5.– С. 29–36 [http://www.viniti.ru/cgi-bin/nti/nti.pl?action=show&year=2\_2002&issue=5&page=29].
3. Блюменау, Д.И. Индикаторный метод компьютерного свертывания в процессе обучения аналитико-синтетической переработке информации / Д.И. Блюменау, Л.Н. Афанасова  // Научные и технические библиотеки.‑ 2001.‑ № 12 [http://www.gpntb.ru/win/ntb/ntb2001/12/f12\_03.htm].
4. Гендина, Н.И. Методика формализованного составления справочной аннотации [Текст] / Н.И. Гендина, И.Л. Скипор //Основы информационной культуры.– Кемерово, 1999.– С. 44‑60. [http://taoi.kemguki.ru/files/%CC%E5%F2%EE%E4%E8%EA%E0%D1%EF%F0%E0%E2%EE%F7%ED%EE%E9%C0%ED%ED%EE%F2%E0%F6%E8%E8.doc].
5. Пащенко, Н.А. Проблемы автоматизации индексирования и реферирования // Итоги науки и техники. Сер. Информатика / Н.А. Пащенко, Л.В. Кнорина, Т.В. Молчанова и др. – М.: ВИНИТИ, 1983.– Т.7.– С. 7 ‑ 164.
6. Ступин, В.С. Система автоматического реферирования методом симметричного реферирования // Компьютерная лингвистика и интеллектуальные технологии" по материалам международной конференции «Диалог».–2004.– С. 579-591. [http://www.dialog-21.ru/Archive/2004/Stupin.pdf].
7. Хан, У. Системы автоматического реферирования / У. Хан, И. Мани // Открытые системы.– 2000 [http://www.osp.ru/os/2000/12/067.htm].
8. Яцко, В.А. Симметричное реферирование: теоретические основы и методика // НТИ.Сер.2. Информационные процессы и системы.–2002. – № 5. – С.18‑28 [http://www.viniti.ru/cgi-bin/nti/nti.pl?action=show&year=2\_2002&issue=5&page=18].

**/Дополнительная**

1. Берзон, В.Е. Классификация коннекторов и диалоговые системы автоматического реферирования / В.Е. Берзон, А.Б. Брайловский // НТИ. Сер.2. Информационные процессы и системы.–1979.– № 11.– С. 19–23.
2. Блюменау, Д.И. Формализованное реферирование с использованием словесных клише (маркеров) / Д.И. Блюменау, Н.И. Гендина, И.С. Добронравов и др. // НТИ. Сер.2. Информационные процессы и системы.– 1981. – № 2. – С. 16–20.
3. Блюменау, Д.И. Проблемы свертывания научной информации.– Л., 1982.
4. Гендина, Н.И. Методика реферирования: Метод. рек. / КГИК.- Кемерово, 1991. ‑ 100 с.
5. Гончаренко, В.В. Фреймы для распознавания смысла текста / В.В. Гончаренко, , Е.А. Шингарева.– Кишинев, 1984.
6. Карпилович, Т.П. Моделирование процесса смысловой компрессии текста: когнитивно-дискурсивный подход. – Мн.: МГЛУ,2003.
7. Леонов, В.П. Реферирование и аннотирование научно-технической литературы.– Новосибирск, 1986.
8. Леонов, В.П. О методах автоматического реферирования (США, 1958 - 1974 гг.) // НТИ.Сер.2. Информационные процессы и системы.– 1975.– № 6.– С. 16–20.
9. Леонов, В.П. Основные методы автоматического реферирования научно-технической литературы (1958–1977). Итоговый отчет НИР 46/ НИИ-31. Разработка принципов и методов автоматического реферирования научно-технической литературы (на примере электротехники), ЛГИК им. Н.К.Крупской.– Л., 1977.– С. 143–185.
10. Нестеров, П.В. Интерактивная система автоматического реферирования документов на основе словаря стандартных фраз / П.В. Нестеров, Г.Г. Белоногов, Ю.Г. Зеленков и др. // НТИ.Сер.2. Информационные процессы и системы.– 1991.– №7.– С.13– 15.
11. Севбо, И.П. Структура связного текста и автоматизация реферирования. – М., 1969.
12. Скороходько, Э.Ф. Семантические сети и автоматическая обработка текста.– Киев, 1983.
13. Солтон, Дж. Динамические библиотечно-информационные системы. – М.: Мир, 1979.– С. 125–132.
14. Яцко, В.А. Логико-лингвистические проблемы анализа и реферирования научного текста. – Абакан: Изд-во Хакасского гос. ун-та, 1996.

**4.2. Критерии оценок результатов учебной деятельности**

|  |  |
| --- | --- |
| **Оценка** | **Критерии оценки уровня знаний и компетенций студента** |
| ***10 баллов*** | Студент дал исчерпывающий ответ на теоретический вопрос, продемонстрировал свободное владение понятийно-терминологическим аппаратом по теме вопроса билета, верно выполнил практическое задание, объяснив алгоритм его выполнения; успешно ответил на дополнительные вопросы преподавателя по программе дисциплины. |
| ***9 баллов*** | Студент дал полный ответ на теоретический вопрос, продемонстрировал владение понятийно-терминологическим аппаратом по теме вопроса билета, верно выполнил практическое задание, успешно ответил на дополнительные вопросы преподавателя по теме, смежной с вопросом / заданием билета.. |
| ***8 баллов*** | Студент дал полный ответ на теоретический вопрос, но допустил незначительные неточности, которые в дальнейшем смог устранить, отвечая на уточняющие вопросы преподавателя, с незначительными погрешностями выполнил практическое задание, но смог внести необходимые исправления по требованию преподавателя. |
| ***7 баллов*** | Студент дал полный, но с незначительными неточностями, ответ на теоретический вопрос, с незначительными погрешностями выполнил практическое задание, и при этом не смог внести необходимые исправления по требованию преподавателя. |
| ***6 баллов*** | Студент дал неполный, но удовлетворительный ответ на теоретический вопрос, с существенными погрешностями выполнил практическое задание, но смог частично внести необходимые исправления по требованию преподавателя. |
| ***5 баллов*** | Студент дал неполный, но удовлетворительный ответ на теоретический вопрос, с существенными погрешностями выполнил практическое задание, и при этом не смог внести необходимые исправления по требованию преподавателя. |
| ***4 балла*** | Студент дал неполный, но удовлетворительный ответ на теоретический вопрос, не выполнил практическое задание, но показал понимание сути задания, отвечая на наводящие вопросы преподавателя. |
| ***3 балла*** | Студент дал неудовлетворительный ответ на теоретический вопрос, не выполнил практическое задание либо выполнил с существенными погрешностями, которые не смог исправить, несмотря на наводящие вопросы преподавателя. |
| ***2 балла*** | Студент не дал ответа на теоретический вопрос, не выполнил практическое задание либо выполнил с существенными погрешностями, которые не смог исправить, несмотря на наводящие вопросы преподавателя. |
| ***1 балл*** | Студент не дал ответа на теоретический вопрос, не выполнил практическое задание либо отказался отвечать. |

Условием допуска к экзамену является успешное выполнение проекта, выполняемого студентом самостоятельно, во внеаудиторное время.

**4.3. Примерный список вопросов к экзамену**

1. Классы методов АОТ. Этапы автоматического анализа текста.
2. Регулярные выражения как запись представления о коллекции строк.
3. Регулярные выражения: применимость в АОТ.
4. Методы автоматического морфологического анализа: декларативный, процедурный, комбинированный.
5. Разметка по частям речи: методы, применимость в АОТ.
6. Модели синтаксической структуры для автоматического синтаксического анализа.
7. Методы автоматического индексирования: статистические, словарные, морфологические, синтаксические.
8. Методы автоматического индексирования: семантические, суперсинтаксические, позиционные.
9. Метод концептуального индексирования (лингвистическая онтология для АИ).
10. Методы извлечения терминов из текста.
11. Архитектура системы автореферирования.
12. Автореферирование и квазиреферирование. Полуавтоматическое реферирование.
13. Методы автоматического реферирования: статистические, позиционные, суперсинтаксические, семантико-синтаксические.
14. Метод симметричного реферирования (В.Яцко, В.Ступин).
15. Метод маркерного реферирования (Д.И. Блюменау и др.) Аспектная сетка документа.
16. Мера TF\*IDF и ее применение в АОТ.
17. Формулы критериев связности: применимость в АОТ, программные средства расчета.
18. Инструментальные средства АОТ: автоматический морфологический анализ.
19. Инструментальные средства АОТ: автоматический синтаксический анализ.
20. Инструментальные средства АОТ: автоматическое извлечение терминов из текста.

**4.4. Примерная тематика проектов**

1. Банк регулярных выражений для извлечения лингвистических объектов из электронного текста.
2. Разработка фрагмента машинного словаря для автоматического морфологического анализа.
3. Разработка фрагмента машинного словаря для автоматического индексирования.
4. Разработка фрагмента лингвистической онтологии для автоматического индексирования.
5. Разработка фрагмента машинного словаря для системы индикаторного реферирования (на материале белорусского языка).
6. Разработка фрагмента семантической сети для заданной тематической группы слов.
7. Сравнительный обзор подборки свободно доступных инструментов АОТ (по видам АОТ).

**СОДЕРЖАНИЕ**

**Пояснительная записка**...…………………………………….….………………………3

**Примерный тематический план** ….…………………….……………………………...5

**Содержание учебного материала**………………………………………….....................6

**Информационно-методическая часть**……………………………………..……...........8

**Критерии оценок результатов учебной деятельности...**………………………..…...15

**Примерный список вопросов к экзамену** ………....……………………………….....16

**Примерная тематика проектов**…………..…………………………………………….17

**Сведения об авторах:**

1. Рычкова Людмила Васильевна – заведующий кафедрой общего и славянского языкознания учреждения образования «Гродненский государственный университет имени Янки Купалы», кандидат филологических наук, доцент.   
Тел. +37529 8895828 (МТС).

2. Станкевич Алеся Юрьевна – старший преподаватель кафедры общего и славянского языкознания учреждения образования «Гродненский государственный университет имени Янки Купалы».