

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Методические указания
по выполнению практических по дисциплине
«Практикум по применению систем автоматизированного анализа текста»

Направление подготовки	45.04.02 Лингвистика
Направленность (профиль)	Современные методы прикладной лингвистики и перевода
Год начала обучения	2026
Форма обучения	очная
Реализуется в семестре	2

Ставрополь, 2026

Введение

Методические рекомендации по выполнению практических работ по дисциплине «Практикум по применению систем автоматизированного анализа текста» разработаны в соответствии с рабочей программой дисциплины по направлению 45.04.02 Лингвистика, профиль – Современные методы прикладной лингвистики и перевода.

Практические задания разработаны в соответствии с рабочей программой дисциплины «Практикум по применению систем автоматизированного анализа текста», целью которой является систематическое изложение основной проблематики формализации языковых явлений, а также развитие практических навыков и умений студентов в части построения языковых моделей и их использования для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Целью практических занятий является закрепление теоретических знаний и приобретение практических умений и навыков, необходимых для освоения принципов и приемов формализации языковых явлений.

Методические рекомендации по каждой практической работе имеют теоретическую часть, необходимую для выполнения практических заданий. Практические задания органично сочетаются с теоретическими знаниями.

Практическое занятие 1.

Тема: Общие принципы применения автоматизированных систем анализа текста.

Регулярные выражения.

Цель: овладение студентами специфическими практическими навыками и умениями, необходимыми для формализации языковых явлений и построения их формальных моделей, а также использования данных моделей для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Реализуемые компетенции:

Индекс	Формулировка:
ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики

Актуальность: актуальность изучения темы определяется существованием практической необходимости в подготовке специалистов в области построения формализованных моделей языковых явлений, что обусловлено существенным поворотом в современной лингвистике в сторону использования компьютерных технологий для обработки и анализа текста по формальным признакам.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ СОБЕСЕДОВАНИЯ

1) Изучите следующий формат описания синтагматических конструкций.

№ №	Условные обозначения	Значения
1	2	3
1	*	Соответствует любому количеству произвольных лексических единиц и знаков препинания кроме терминальных (завершающих предложение)
2	\\	Соответствует границе предложения. Если предложение содержит прямую речь, то прямая речь выделяется как отдельное предложение или совокупность предложений
3	[*]	Соответствует одной произвольной лексической единице
4	[элемент]	Соответствует приведенному в скобках слову, нетерминальному знаку препинания или словосочетанию. Указывается буквами в нижнем регистре, например: [like]
5	[ГРУППА]	Соответствует группе слов, приведенной в квадратных скобках. Указывается буквами в верхнем регистре, например: [PERSONAL_PRONOUN] – соответствует одному любому личному местоимению
6	<условие>	Условие, указанное в угловых скобках, является необязательным для выполнения, например: <> - соответствует наличию или отсутствию запятой
7	{эл.1 эл.2 ... эл.N}	Соответствует одному любому из приведенных в фигурных скобках равновероятных элементов, например: {you know y'know}
8	(знак)	Соответствует указанному в скобках терминальному знаку препинания, например: (.)
9	(зн.1 зн.2 ... зн.N)	Соответствует одному из указанных терминальных знаков препинания, например: (. ...!)

Под терминальными понимаются знаки препинания, используемые для завершения предложения (точка, восклицательный знак, вопросительный знак, многоточие, а также их комбинации). К нетерминальным, напротив, относятся знаки препинания, которые не могут быть использованы для завершения предложения (запятая, двоеточие, дефис и т.п.).

2) Распишите следующие формализованные представления регулярных выражений с помощью метасимволов Perl.

```
\\ [why] [,] * (.) \\
```

Данное регулярное выражение трактуется следующим образом: в начале предложения должно находиться слово *why*, отделенное запятой. За ним может следовать любое количество произвольных лексических единиц и нетерминальных знаков препинания, после чего предложение должно заканчиваться точкой.

Важным представляется то, что слова и словосочетания в квадратных скобках всегда указываются со строчной буквы. Это связано с тем, что автоматизированный лексический анализатор текста перед сравнением предложения с шаблоном приводит все слова к нижнему регистру, чтобы обеспечить независимость хода анализа от регистра, в котором написаны слова в предложении. В случае, если регистр, в котором написано то или иное слово, может повлиять на результат анализа, на этапе предварительной обработки предложения проводится дополнительный, чувствительный к регистру символов анализ.

Итак, регулярному выражению 1) будут соответствовать, например, следующие предложения:

а) *Why*, I think it's quite obvious. (Из разг. речи).

б) *Why*, it's very interesting. (Из разг. речи).

Предложение «*Why*, of course!», напротив, не соответствует регулярному выражению 1), так как терминальный знак препинания в данном случае – восклицательный знак, а не точка.

В случае, если на определенном месте в предложении может находиться одна из нескольких равновероятных лексических единиц, что не влияет на результат анализа, используется условное выражение вида «{слово|слово|...}», например:

```
\\ [I] [,] {uh|um} * (...)
```

Регулярному выражению 2) будут соответствовать следующие предложения:

а) *I, uh*... don't even know where to start. (Из разг. речи).

б) *I, um*... I'm not interested in computer animation. (Из разг. речи).

Список рекомендуемой литературы

Основная литература

1. Баймуратова, У. Электронный инструментарий переводчика : учебное пособие / У. Баймуратова ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург : ОГУ, 2013. - 120 с. ; То же [Электронный ресурс]. - URL:<http://biblioclub.ru/index.php?page=book&id=259202>

2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

Дополнительная литература

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.

3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>

4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O’Reilly Media, 2012. – 544 p.

Методическая литература

1. Методические рекомендации к самостоятельной работе студентов по дисциплине «Практикум по применению систем автоматизированного анализа текста». Каменский М.В. – Ставрополь, 2026

Интернет-ресурсы

1. Applied Linguistics – <http://www.appliedlinguistics.org>

2. Center for Applied Linguistics – <http://www.cal.org>

3. GATE (Sheffield University, UK) – <http://gate.ac.uk>

Программное обеспечение

1. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>

2. Applied Linguistics – <http://www.appliedlinguistics.org>

3. Center for Applied Linguistics – <http://www.cal.org>

4. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>

5. Emacs – <http://www.gnu.org/software/emacs>

6. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>

7. GNU Compiler Collection – <http://gcc.gnu.org>

8. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>

9. Leopard Language Parser – <http://leopard.loria.fr>

10. NLTK (Natural Language Toolkit) – <http://www.nltk.org>

11. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>

12. Praat – <http://www.fon.hum.uva.nl/praat>

13. Python Programming Language – <http://www.python.org>

14. Sed – <http://www.gnu.org/s/sed>

15. Sonic Visualizer – <http://www.sonicvisualiser.org>

16. Vi Improved - <http://www.vim.org>

Практическое занятие 2.

Тема: Метасимволы регулярных выражений (I)

Цель: овладение студентами специфическими практическими навыками и умениями, необходимыми для формализации языковых явлений и построения их формальных моделей, а также использования данных моделей для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Реализуемые компетенции:

ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики
------	--

Актуальность: актуальность изучения темы определяется существованием практической необходимости в подготовке специалистов в области построения формализованных моделей языковых явлений, что обусловлено существенным поворотом в современной лингвистике в сторону использования компьютерных технологий для обработки и анализа текста по формальным признакам.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ СОБЕСЕДОВАНИЯ

Распишите следующие формализованные представления регулярных выражений с помощью метасимволов Perl.

1. Дискурсный маркер *well*.

Дискурсный маркер *well* может быть в большинстве случаев идентифицирован по занимаемой им начальной позиции в предложении. Как правило, данный дискурсный маркер отделяется запятой, поскольку входит в отдельную интонационную группу.

В случае, если дискурсный маркер *well* находится в повествовательном предложении, то он может использоваться либо с целью смены темы разговора, либо для выражения собственной точки зрения, либо для смягчения высказывания. В данном случае на основную функцию дискурсного маркера может также наслаиваться фокусная функция. В общем виде регулярное выражение, описывающее дискурсный маркер *well* в приведенных функциях, может быть представлено в следующем виде:

1. `\\ [well] [,] * (.) \\`

Например:

а) *Well*, that's no big deal for me, I don't really care. (Из разг. речи).

б) *Well*, let's speak about your recent promotion! (Из разг. речи).

в) *Well*, as I was saying before, it's not too important. (Из разг. речи).

Стоит отметить, что при выражении собственной точки зрения дискурсный маркер *well* может сочетаться с дискурсным маркером *why*, с которым он часто входит в одну интонационную группу, при этом дискурсный маркер *why* может, но не обязательно, быть отделен запятой. В данном случае для распознавания дискурсного маркера *well* применяется совокупность регулярных выражений вида:

2. `\\ [well] [why] [,] * \\`

3. `\\ [well] [,] [why] [,] * \\`

Например:

а) A: Do you think we could just go inside and grab the book?

B: *Well why*, of course, it's going to be really easy! (Из разг. речи).

б) A: I don't think it's an easy task.

B: *Well, why*, on the contrary – it's a piece of cake! (Из разг. речи).

В случае, если дискурсный маркер *well* используется с целью смягчения отрицательного ответа на вопрос, то до предложения с дискурсным маркером, как правило, следует вопросительное предложение, что может быть выражено с помощью регулярного выражения следующего вида:

4. \\ * (?) \\ [well] [,] * (.) \\

Например:

а) A: Can I borrow your textbook?

B: *Well*, I need it tonight in order to study. (Из разг. речи).

б) A: Do you have any classes tomorrow?

B: *Well*, yes, but nothing too important. (Из разг. речи).

В случае, если *well* используется для выражения недовольства или побуждения к действию, то он употребляется в восклицательном предложении, и определяется по регулярному выражению вида:

5. \\ [well] [,] * (!) \\

Например:

а) *Well*, I don't think you're supposed to be here, go away! (Из разг. речи).

б) *Well*, I told you that you had to be careful! (Из разг. Речи).

Список рекомендуемой литературы

Основная литература

1. Баймуратова, У. Электронный инструментарий переводчика : учебное пособие / У. Баймуратова ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург : ОГУ, 2013. - 120 с. ; То же [Электронный ресурс]. - URL:<http://biblioclub.ru/index.php?page=book&id=259202>

2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

Дополнительная литература

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.

3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>

4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O'Reilly Media, 2012. – 544 p.

Методическая литература

1. Методические рекомендации к самостоятельной работе студентов по дисциплине «Практикум по применению систем автоматизированного анализа текста». Каменский М.В. – Ставрополь, 2026

Интернет-ресурсы

1. Applied Linguistics – <http://www.appliedlinguistics.org>

2. Center for Applied Linguistics – <http://www.cal.org>

3. GATE (Sheffield University, UK) – <http://gate.ac.uk>

Программное обеспечение

1. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>

2. Applied Linguistics – <http://www.appliedlinguistics.org>

3. Center for Applied Linguistics – <http://www.cal.org>

4. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>

5. Emacs – <http://www.gnu.org/software/emacs>
6. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>
7. GNU Compiler Collection – <http://gcc.gnu.org>
8. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>
9. Leopard Language Parser – <http://leopard.loria.fr>
10. NLTK (Natural Language Toolkit) – <http://www.nltk.org>
11. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>
12. Praat – <http://www.fon.hum.uva.nl/praat>
13. Python Programming Language – <http://www.python.org>
14. Sed – <http://www.gnu.org/s/sed>
15. Sonic Visualizer – <http://www.sonicvisualiser.org>
- 16. Vi Improved - <http://www.vim.org>**

Практическое занятие 3.

Тема: Метасимволы регулярных выражений (II)

Цель: овладение студентами специфическими практическими навыками и умениями, необходимыми для формализации языковых явлений и построения их формальных моделей, а также использования данных моделей для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Реализуемые компетенции:

ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики
------	--

Актуальность: актуальность изучения темы определяется существованием практической необходимости в подготовке специалистов в области построения формализованных моделей языковых явлений, что обусловлено существенным поворотом в современной лингвистике в сторону использования компьютерных технологий для обработки и анализа текста по формальным признакам.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ СОБЕСЕДОВАНИЯ

Изучите нижеприведенные языковые модели и распишите их с применением метасимволов Perl.

Если дискурсный маркер *well* применяется как маркер планирования с целью обдумывания высказывания или ответа на вопрос, то за дискурсным маркером следует многоточие, при этом между дискурсным маркером и многоточием могут находиться несколько дополнительных лексических единиц. Стоит отметить, что при употреблении в данной функции дискурсный маркер может находиться в любой части предложения. Таким образом, для идентификации дискурсного маркера *well* в данной функции целесообразно использование следующей совокупности регулярных выражений, первое из которых является более точным, чем второе:

6. `\\ * [well] (...)`

7. `\\ * [well] * (...)`

Например:

а) *Well...* I don't know, I'm not sure... (Из разг. речи).

б) *Well I...* I don't know, I'm not sure... (Из разг. речи).

в) I want to say... *well...*, *well I...* No, I can't say this! (Из разг. речи).

Использование двух регулярных выражений диктуется необходимостью достижения максимально возможной точности определения дискурсных маркеров в тексте. Таким образом, поиск по регулярному выражению 7) осуществляется только в случае, когда предложение не соответствует регулярному выражению 6), то есть предложение не совпадает с более точным регулярным выражением.

В случае, если слово *well* находится в начальной позиции в предложении, но предложение не соответствует регулярным выражениям 1) – 7) и, таким образом, однозначно идентифицировать дискурсивный статус слова *well* не представляется возможным, подобное употребление слова *well* может быть отслежено по желанию пользователя с целью дальнейшего исследования. Таким образом, после проверок на соответствие предложения регулярным выражениям 1) – 7) выполняется проверка на соответствие регулярному выражению вида:

8. `\\ [well] * \\`

Список рекомендуемой литературы

Основная литература

1. Баймуратова, У. Электронный инструментарий переводчика : учебное пособие / У. Баймуратова ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург : ОГУ, 2013. - 120 с. ; То же [Электронный ресурс]. - URL:<http://biblioclub.ru/index.php?page=book&id=259202>

2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

Дополнительная литература

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.

3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>

4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O'Reilly Media, 2012. – 544 p.

Методическая литература

1. Методические рекомендации к самостоятельной работе студентов по дисциплине «Практикум по применению систем автоматизированного анализа текста». Каменский М.В. – Ставрополь, 2026

Интернет-ресурсы

1. Applied Linguistics – <http://www.appliedlinguistics.org>

2. Center for Applied Linguistics – <http://www.cal.org>

3. GATE (Sheffield University, UK) – <http://gate.ac.uk>

Программное обеспечение

1. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>

2. Applied Linguistics – <http://www.appliedlinguistics.org>

3. Center for Applied Linguistics – <http://www.cal.org>

4. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>

5. Emacs – <http://www.gnu.org/software/emacs>

6. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>

7. GNU Compiler Collection – <http://gcc.gnu.org>

8. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>

9. Leopard Language Parser – <http://leopard.loria.fr>

10. NLTK (Natural Language Toolkit) – <http://www.nltk.org>

11. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>

12. Praat – <http://www.fon.hum.uva.nl/praat>

13. Python Programming Language – <http://www.python.org>

14. Sed – <http://www.gnu.org/s/sed>

15. Sonic Visualizer – <http://www.sonicvisualiser.org>

16. Vi Improved - <http://www.vim.org>

Практическое занятие 4.

Тема: Четкий и нечеткий поиск с применением регулярных выражений (I)

Цель: овладение студентами специфическими практическими навыками и умениями, необходимыми для формализации языковых явлений и построения их формальных моделей, а также использования данных моделей для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Реализуемые компетенции:

ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики
------	--

Актуальность: актуальность изучения темы определяется существованием практической необходимости в подготовке специалистов в области построения формализованных моделей языковых явлений, что обусловлено существенным поворотом в современной лингвистике в сторону использования компьютерных технологий для обработки и анализа текста по формальным признакам.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ СОБЕСЕДОВАНИЯ

Изучите нижеприведенные языковые модели и распишите их с применением метасимволов Perl.

2. Дискурсный маркер *oh*.

Слово *oh* независимо от позиции в предложении является дискурсным маркером, что облегчает задачу его автоматизированного распознавания. В общем случае дискурсный маркер *oh* распознается по регулярному выражению вида:

- $\backslash \backslash * [oh] * \backslash \backslash$

В случае, если дискурсный маркер *oh* используется с целью указания на неточность приводимой информации или на ее приблизительный характер, или же употребляется с целью коррекции предыдущего высказывания, маркер может находиться между уточняемой и уточненной частями предложения и либо выделяется запятой с двух сторон, либо находится за многоточием и выделяется запятой спереди. Предложение с дискурсным маркером *oh* в указанной функции может быть повествовательным или восклицательным, при этом семантическая завершенность или незавершенность предложения не влияет на функцию дискурсного маркера. Регулярные выражения, описывающие приведенные ситуации, могут быть представлены в виде:

- $\backslash \backslash [*] * [,] [oh] [,] * (|\dots|!) \backslash \backslash$
- $\backslash \backslash * (\dots) \backslash \backslash [*] [oh] [,] * (|\dots|!) \backslash \backslash$

Например:

а) I think it was four days ago, *oh*, actually... maybe it was three days ago. (Из разг. речи).

б) I think it was in 1972... *oh*, no, probably in 1973. (Из разг. речи).

Если дискурсный маркер *oh* находится в начале предложения, он также отделяется запятой. Функция маркера *oh* зависит от типа предложения. Так, если предложение является повествовательным или восклицательным, то маркер может быть использован либо для введения уточненной или скорректированной информации, либо для обеспечения обратной связи. Данная ситуация описывается с помощью регулярного выражения вида:

- $\backslash \backslash [oh] [,] * (|\dots|!) \backslash \backslash$

Например:

а) I'm quite sure it was Neal who passed the test with the best score. *Oh*, wait... no, it was actually Alex. (Из разг. речи).

б) A: Actually, Dorian was present at Jack's birthday party – Jack finally decided to invite him!

B: *Oh*, that's interesting news! (Из разг. речи).

В противном случае, если предложение является вопросительным, то дискурсный маркер *oh* используется либо для запроса дополнительной информации коммуникантом, либо для фокусирования внимания реципиента на предложении, либо сочетает обе эти функции:

- \ [oh] [,] * (?) \

Например:

а) *Oh*, do you know what this book is about? (Из разг. речи).

б) *Oh*, did you know that Ellen has written her new poem today? (Из разг. речи).

Список рекомендуемой литературы

Основная литература

1. Баймуратова, У. Электронный инструментарий переводчика : учебное пособие / У. Баймуратова ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург : ОГУ, 2013. - 120 с. ; То же [Электронный ресурс]. - URL:<http://biblioclub.ru/index.php?page=book&id=259202>
2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

Дополнительная литература

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.
3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>
4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O’Reilly Media, 2012. – 544 p.

Методическая литература

1. Методические рекомендации к самостоятельной работе студентов по дисциплине «Практикум по применению систем автоматизированного анализа текста». Каменский М.В. – Ставрополь, 2026

Интернет-ресурсы

1. Applied Linguistics – <http://www.appliedlinguistics.org>
2. Center for Applied Linguistics – <http://www.cal.org>
3. GATE (Sheffield University, UK) – <http://gate.ac.uk>

Программное обеспечение

1. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>
2. Applied Linguistics – <http://www.appliedlinguistics.org>
3. Center for Applied Linguistics – <http://www.cal.org>
4. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>
5. Emacs – <http://www.gnu.org/software/emacs>
6. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>
7. GNU Compiler Collection – <http://gcc.gnu.org>
8. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>
9. Leopard Language Parser – <http://leopard.loria.fr>

10. NLTK (Natural Language Toolkit) – <http://www.nltk.org>
11. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>
12. Praat – <http://www.fon.hum.uva.nl/praat>
13. Python Programming Language – <http://www.python.org>
14. Sed – <http://www.gnu.org/s/sed>
15. Sonic Visualizer – <http://www.sonicvisualiser.org>
16. **Vi Improved - <http://www.vim.org>**

Практическое занятие 5.

Тема: Четкий и нечеткий поиск с применением регулярных выражений (II).

Цель: овладение студентами специфическими практическими навыками и умениями, необходимыми для формализации языковых явлений и построения их формальных моделей, а также использования данных моделей для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Реализуемые компетенции:

ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики
------	--

Актуальность: актуальность изучения темы определяется существованием практической необходимости в подготовке специалистов в области построения формализованных моделей языковых явлений, что обусловлено существенным поворотом в современной лингвистике в сторону использования компьютерных технологий для обработки и анализа текста по формальным признакам.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ СОБЕСЕДОВАНИЯ

Изучите нижеприведенные языковые модели и распишите их с применением метасимволов Perl.

3. Дискурсный маркер *you know*.

Данный дискурсный маркер хорошо поддается описанию с помощью регулярных выражений, поскольку занимает либо начальную, либо конечную позицию в предложении и, как правило, выделяется на письме запятой. Как было показано в разделе 2.2.1, дискурсный маркер *you know* имеет две различные формы произношения и написания – *you know* и *y'know*, что указывает на необходимость учета во всех описывающих его регулярных выражениях условия вида «*{you know|y'know}*», позволяющего принять во внимание обе эти формы.

В случае, когда дискурсный маркер *you know* используется как эмфатический маркер или как маркер, указывающий реципиенту на известность или очевидность информации в предложении, он может занимать начальную или конечную позицию в предложении и отделяется запятой. Предложение в данном случае может заканчиваться либо точкой, либо многоточием, восклицательные же предложения имеет смысл вынести в отдельную группу в силу того, что дискурсный маркер *you know* может употребляться в них в иных функциях.

В начальной позиции возможна ситуация, в которой дискурсный маркер является составляющей частью выражения *you know what*, что облегчает его автоматизированное распознавание.

Таким образом, для распознавания маркера *you know* в указанных функциях используется совокупность регулярных выражений вида:

- `\\ {you know what|y'know what} [,] * (\\.\\.\\.)\ \`
- `\\ {you know|y'know} [,] * (\\.\\.\\.)\ \`
- `\\ * [,] {you know|y'know} (\\.\\.\\.)\ \`

Регулярные выражения 1) и 2) служат для распознавания дискурсного маркера в начале предложения, а выражение 3) – в конце. Например:

а) *You know*, it's quite obvious that you're supposed to do your homework daily. (Из разг. речи).

б) *Y'know*, Jack is a real fool... (Из разг. речи).

в) It's the most modern personal computer today, *you know*. (Из разг. речи).

Также в случае, когда дискурсный маркер *you know* находится в начале предложения, ему может предшествовать другой дискурсный маркер, чаще всего *well* в функции выражения точки зрения говорящего, выражения недовольства или же в фокусной функции. Для того, чтобы корректно распознать маркер *you know* в подобных предложениях, необходимо наличие дополнительных версий регулярных выражений 1) и 2) с менее жестким условием поиска,

допускающих наличие дополнительных лексических единиц и знаков препинания перед дискурсным маркером *you know*:

- \\ * [,] {you know what|y'know what} [,] * (.|...)\
- \\ * [,] {you know|y'know} [,] * (.|...)\
- \\ * {you know what|y'know what} [,] * (.|...)\
- \\ * {you know|y'know} [,] * (.|...)\

Проверка на соответствие предложения данным регулярным выражениям выполняется только в случае, если не установлено соответствие предложения выражениям 1) и 2). Регулярные выражения 4) и 5) требуют наличия перед дискурсным маркером *you know* запятой и потому являются более точными и приоритетными по отношению к выражениям 6) и 7), которые позволяют только отследить факт наличия в повествовательном предложении конструкции «{*you know|y'know*}», отделенной запятой. Выражения 6) и 7) используются для выделения случаев, в которых выражение *you know* предположительно является дискурсным маркером, при этом соответствие предложения более точным регулярным выражениям 1),2),4),5) не установлено. Например:

- a) Well, *you know*, it's an excellent opportunity to show what I can do! (Из разг. речи).
- б) Well *you know*, my trip to China was simply brilliant! (Из разг. речи).

В случае, если дискурсный маркер *you know* используется как эмфатический маркер с целью указания на значимость высказывания, он часто используется в восклицательных предложениях и может занимать как начальную, так и конечную позицию в предложении, при этом он также отделяется запятой. Таким образом, для его распознавания используются регулярные выражения вида:

- \\ {you know|y'know} [,] * (!)\
- \\ * [,] {you know|y'know} (!)\

Например:

- a) *You know*, it's not a joke! (Из разг. речи).
- б) You're taking this too easily, but I'm talking about serious things, *y'know*! (Из разг. речи).

Как и в предыдущем случае, с целью учета случаев употребления дискурсного маркера *you know* в начале предложения совместно с другими дискурсными маркерами используется дополнительное регулярное выражение, менее приоритетное по отношению к выражению 8):

- \\ * {you know|y'know} [,] * (!)\

Например:

- a) Well, *you know*, you have to treat your new equipment with care! (Из разг. речи).
- б) Well *you know*, this is not the point! (Из разг. речи).

Дискурсный маркер *you know* в качестве маркера обратной связи применяется в вопросительных предложениях и входит в выражение *you know what*. Чаще всего встречается либо изолированная конструкция *you know what* (то есть в предложении нет других лексических единиц), либо конструкция *you know what*, находящаяся в конечной позиции в предложении и отделенная запятой. Таким образом, наиболее приоритетными регулярными выражениями для поиска дискурсного маркера *well* в данной функции будут следующие:

- \\ {you know what|y'know what} (?)\
- \\ * [,] <[*]> {you know what|y'know what} (?)\

Например:

- a) A: *You know what*?
B: Yeah? (Из разг. речи).
- б) A: I thought about your offer all these days, but *you know what*?
B: What?
A: I decided to say no. (Из разг. речи).

В качестве дополнительного средства нахождения конструкции *you know what* в вопросительных предложениях применяется регулярное выражение вида:

13) \\ * {you know what|y'know what} * (?) \\

Список рекомендуемой литературы

Основная литература

1. Баймуратова, У. Электронный инструментарий переводчика : учебное пособие / У. Баймуратова ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург : ОГУ, 2013. - 120 с. ; То же [Электронный ресурс]. - URL:<http://biblioclub.ru/index.php?page=book&id=259202>
2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

Дополнительная литература

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.
3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>
4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O'Reilly Media, 2012. – 544 p.

Методическая литература

1. Методические рекомендации к самостоятельной работе студентов по дисциплине «Практикум по применению систем автоматизированного анализа текста». Каменский М.В. – Ставрополь, 2026

Интернет-ресурсы

1. Applied Linguistics – <http://www.appliedlinguistics.org>
2. Center for Applied Linguistics – <http://www.cal.org>
3. GATE (Sheffield University, UK) – <http://gate.ac.uk>

Программное обеспечение

1. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>
2. Applied Linguistics – <http://www.appliedlinguistics.org>
3. Center for Applied Linguistics – <http://www.cal.org>
4. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>
5. Emacs – <http://www.gnu.org/software/emacs>
6. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>
7. GNU Compiler Collection – <http://gcc.gnu.org>
8. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>
9. Leopard Language Parser – <http://leopard.loria.fr>
10. NLTK (Natural Language Toolkit) – <http://www.nltk.org>
11. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>
12. Praat – <http://www.fon.hum.uva.nl/praat>
13. Python Programming Language – <http://www.python.org>
14. Sed – <http://www.gnu.org/s/sed>
15. Sonic Visualizer – <http://www.sonicvisualiser.org>
16. Vi Improved - <http://www.vim.org>

Практическое занятие 6.

Тема: Четкий и нечеткий поиск с применением регулярных выражений (III)

Цель: овладение студентами специфическими практическими навыками и умениями, необходимыми для формализации языковых явлений и построения их формальных моделей, а также использования данных моделей для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Реализуемые компетенции:

ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики
------	--

Актуальность: актуальность изучения темы определяется существованием практической необходимости в подготовке специалистов в области построения формализованных моделей языковых явлений, что обусловлено существенным поворотом в современной лингвистике в сторону использования компьютерных технологий для обработки и анализа текста по формальным признакам.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ СОБЕСЕДОВАНИЯ

Изучите нижеприведенные языковые модели и распишите их с применением метасимволов Perl.

4. Дискурсный маркер *I mean*.

Словосочетание *I mean* является дискурсным маркером в одном из двух случаев: 1) если располагается в самом начале предложения и выделено запятой, при этом дискурсный маркер *I mean* используется для указания на то, что в предложении перефразировано или пояснено предыдущее высказывание; 2) если располагается между двумя высказываниями в предложении, при этом второе высказывание перефразирует или поясняет первое. Во втором случае дискурсный маркер *I mean*, как правило, отделяется с двух сторон запятой. Таким образом, синтагматику дискурсного маркера *I mean* в указанных случаях удобно представить в виде следующих регулярных выражений:

- $\backslash [i\ mean] [,] * \backslash$
- $\backslash * [,] [i\ mean] [,] * \backslash$

Например, следующие предложения будут соответствовать приведенным регулярным выражениям:

- a) No!... *I mean*, no, he is not a disbeliever, and neither am I. (Из разг. речи).
- б) Don't sell this for money, *I mean*, understandably, you could, since it's basically the size of a novella, but please don't do it. (Из разг. речи).

Стоит отметить, что достаточно часто можно видеть предложения, в которых вместо запятой с целью отделения дискурсного маркера *I mean* использован дефис. Таким образом, с целью повышения точности поиска необходимо уточнить вышеприведенные регулярные выражения следующим образом, включив в них возможность использования дефиса:

- $\backslash [i\ mean] \{,-\} * \backslash$
- $\backslash * \{,-\} [i\ mean] \{,-\} * \backslash$

Таким образом, такие варианты примера б), как «Don't sell this for money, *I mean* – understandably, you could...», «Don't sell this for money – *I mean*, understandably, you could...» и т.п. будут соответствовать заданным критериям поиска дискурсного маркера *I mean*.

Также не исключены варианты пунктуации, когда после дискурсного маркера *I mean* следует многоточие, указывающее на то, что говорящему потребовалось дополнительное время для подготовки перефразированного высказывания. Для распознавания указанных ситуаций используются следующие регулярные выражения:

- $\backslash [i\ mean] (\dots) \backslash$
- $\backslash * [,] [i\ mean] (\dots) \backslash$

Например:

а) It's going to be easy to pass this exam, *I mean*... if you prepare well. (Из разг. речи).

б) This new dual-core processor is twice as fast. *I mean*... if you run software that's specifically designed for it, that is. (Из разг. речи).

В качестве дополнительной опции поиска, позволяющей увеличить точность нахождения дискурсного маркера *I mean*, используется поиск по регулярному выражению вида:

- `\\ [i mean] * \\`

Поскольку данное регулярное выражение является наиболее общим по отношению к предыдущим и позволяет указать лишь вероятность употребления конструкции *I mean* в качестве дискурсного маркера, то проверка на соответствие предложения данному регулярному выражению должна выполняться только для тех предложений, для которых условия 1) – 6) не выполняются. Регулярное выражение 7) позволяет найти дискурсные маркеры *I mean* в случае, если пунктуационные правила не соблюдены, например:

This new dual-core processor is twice as fast. *I mean* if you run software that's specifically designed for it, that is. (Из разг. Речи).

Список рекомендуемой литературы

Основная литература

1. Баймуратова, У. Электронный инструментарий переводчика : учебное пособие / У. Баймуратова ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург : ОГУ, 2013. - 120 с. ; То же [Электронный ресурс]. - URL:<http://biblioclub.ru/index.php?page=book&id=259202>

2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

Дополнительная литература

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.

3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>

4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O'Reilly Media, 2012. – 544 p.

Методическая литература

1. Методические рекомендации к самостоятельной работе студентов по дисциплине «Практикум по применению систем автоматизированного анализа текста». Каменский М.В. – Ставрополь, 2026

Интернет-ресурсы

1. Applied Linguistics – <http://www.appliedlinguistics.org>
2. Center for Applied Linguistics – <http://www.cal.org>
3. GATE (Sheffield University, UK) – <http://gate.ac.uk>

Программное обеспечение

1. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>
2. Applied Linguistics – <http://www.appliedlinguistics.org>
3. Center for Applied Linguistics – <http://www.cal.org>

4. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>
5. Emacs – <http://www.gnu.org/software/emacs>
6. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>
7. GNU Compiler Collection – <http://gcc.gnu.org>
8. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>
9. Leopard Language Parser – <http://leopard.loria.fr>
10. NLTK (Natural Language Toolkit) – <http://www.nltk.org>
11. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>
12. Praat – <http://www.fon.hum.uva.nl/praat>
13. Python Programming Language – <http://www.python.org>
14. Sed – <http://www.gnu.org/s/sed>
15. Sonic Visualizer – <http://www.sonicvisualiser.org>
16. **Vi Improved - <http://www.vim.org>**

Практическое занятие 7.

Тема: Регулярные выражения в Vim.

Цель: овладение студентами специфическими практическими навыками и умениями, необходимыми для формализации языковых явлений и построения их формальных моделей, а также использования данных моделей для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Реализуемые компетенции:

ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики
------	--

Актуальность: актуальность изучения темы определяется существованием практической необходимости в подготовке специалистов в области построения формализованных моделей языковых явлений, что обусловлено существенным поворотом в современной лингвистике в сторону использования компьютерных технологий для обработки и анализа текста по формальным признакам.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ СОБЕСЕДОВАНИЯ

Изучите нижеприведенные языковые модели и распишите их с применением метасимволов Vim.

5. Дискурсный маркер *like*.

В силу того, что в ходе исторического развития английского языка дискурсный маркер *like* в значительной степени подвергся явлениям десемантизации и деграмматизации, он имеет ряд особенностей, усложняющих его автоматизированное распознавание. Во-первых, данный маркер может занимать произвольную позицию в предложении и не всегда выделяется на письме знаками препинания. Во-вторых, он часто сочетает различные функции, которые могут быть распознаны только с учетом семантики предложения и контекста. В-третьих, отдельные функции данного дискурсного маркера (смягчения высказывания, введения необычного описания и выражения незначительного различия между имплицатурой и эксплицатурой) детерминированы в большей степени семантикой высказывания и не могут быть автоматически распознаны исходя исключительно из синтагматических признаков.

Несмотря на указанные особенности, автоматизированное распознавание дискурсного маркера *like* в большинстве его функций представляется возможным. Случаи же употребления *like* как дискурсного маркера, не пригодные для автоматизированного распознавания в силу недостаточной синтаксической информации, могут быть учтены с помощью автоматизированного поиска слова *like* с возможностью быстрого просмотра найденных предложений в контексте.

В случае, если дискурсный маркер *like* используется в функции фокуса предложения, то часто входит в конструкцию «*it's like*» или «*and it's like*», отделяемую запятой. В свою очередь, элемент «*it's*» может быть представлен либо собственно в виде «*it's*», либо в виде «*it is*». В случае, если употреблена конструкция «*it's like*», ей в большинстве случаев предшествует пауза, и она часто выделяется на письме запятой. Конструкция «*and it's like*» может иногда быть представлена в сокращенном виде «*and like*». Как правило, за всеми указанными конструкциями следует короткая пауза, которая на письме имеет несколько распространенных вариантов выражения и может быть представлена запятой, двоеточием, многоточием или дефисом. Еще одним распространенным вариантом оформления конструкций «*it is like / and it is like / and like*» на письме является отделение слова *like* запятой спереди, что также необходимо учесть при построении регулярных выражений. Таким образом, регулярные выражения для определения дискурсного маркера *like* в указанных случаях могут быть представлены в виде:

- `\\ * {and it is <[,]> like|and it's <[,]> like} {:, | - | ...} * \\`
- `\\ * [,] {it is <[,]> like|it's <[,]> like} {:, | - | ...} * \\`
- `\\ * [and <[,]> like] {:, | - | ...} * \\`

Например:

а) After the ceremony Miles comes to me and it's *like*... he's really impressed or excited by something! (Из разг. речи).

б) I enter the room and, it's *like*, no one's home... (Из разг. речи).

в) And then I walked into the house and *like*, found out that the book was still there, lying on the table. (Из разг. речи).

Важным представляется тот факт, что в примере а) после слова *and* стоит запятая, при этом предложение находится по соответствию регулярному выражению 2), а не 1).

В некоторых случаях конструкция «*it is like*» может не отделяться спереди запятой. Для более точного ее распознавания используется обобщенный вариант регулярного выражения 2):

- `\\ * {it is <[,] like|it's <[,]> like} {:, | - | ...} * \\`

Если дискурсный маркер *like* используется как маркер, сигнализирующий о необходимости дополнительного времени для обдумывания высказывания, за ним, как правило, следует многоточие. Часто многоточие находится и перед дискурсным маркером *like*. Таким образом, регулярные выражения для определения данной функции дискурсного маркера *like* можно записать в виде:

- `\\ * [...] [like] [...] * \\`
- `\\ * [like] [...] * \\`

Например:

а) Well, I... *like*... don't know what to say. (Из разг. речи).

б) As I was saying, *like*... uh... it's definitely a modern way of programming. (Из разг. речи).

Стоит отметить, что регулярные выражения 1) – 4), особенно в случае, если после конструкции с дискурсным маркером *like* находится многоточие, могут также соответствовать предложениям с маркером *like* в функции маркера планирования, при этом функция планирования может наслаиваться на функцию фокуса предложения.

При эмоциональном цитировании дискурсному маркеру *like* часто предшествует местоимение или собственное имя и глагол *to be* в форме третьего лица единственного или множественного числа. Указанная конструкция, как правило, отделяется двоеточием или (реже) запятой. Также допускается многоточие, особенно в случаях, когда цитируется не речь, а мимика другого человека. Таким образом, регулярное выражение для поиска дискурсного маркера *like* в функции эмоционально нагруженного цитирования может быть записано в виде:

- `\\ * {PERSONAL_PRONOUN|PROPER_NOUN} {is|are|was|were|will be} [like] {:, | , | ...} * \\`

Например:

а) And then, when I saw this girl, she was *like*: “Hey, Alex! How's it going?” (Из разг. речи).

б) I was in my office, and then Mr. Walter was *like*: “Start doing something already, will you?!” (Из разг. речи).

В других случаях перед глаголом *to be* и дискурсным маркером *like* может находиться имя нарицательное или союз *and*. В данном случае для отслеживания дискурсного маркера *like* применяется следующий обобщенный вариант регулярного выражения 7), являющийся менее приоритетным в силу меньшей точности распознавания слова *like* как дискурсного маркера:

- `\\ * {is|are|was|were|will be} [like] {:, | , | ...} * \\`

Например:

а) I entered the office and saw a stranger there. The man was *like*: “I was expecting you, Mr. Olson!” (Из разг. речи).

б) I stood before him and was *like*... (далее нет слов, передается только мимика цитируемого человека) (Из разг. речи).

Дискурсный маркер *like*, используемый для указания на неточность информации в предложении, часто сопровождается конструкцией с союзом *or*, при этом дискурсный маркер *like* может отделяться запятой или многоточием сзади. Распознавание маркера *like* в данной функции осложняется тем, что между словом *like* и непосредственно частью предложения, в которой выражается сомнение, может находиться дополнительная лексическая единица, чаще

всего предлог. Лексические единицы, разделенные союзом *or*, могут быть представлены словами, относящимися к различным частям речи, однако наиболее распространены конструкции с числительными (например, *fifty or sixty*).

Таким образом, для распознавания дискурсного маркера *like* в указанной функции применяются следующие регулярные выражения:

- для конструкций с числительными, разделенными союзом «*or*»:

9) $\backslash \backslash * [like] <\{, | \dots\} <[*]> [NUMERAL] <[,]> [or] * \backslash$

- для конструкций со словами, разделенными союзом «*or*», но не являющихся числительными:

10) $\backslash \backslash * [like] <\{, | \dots\} <[*]> [*] <[,]> [or] * \backslash$

Например:

а) I think we got acquainted in *like*, 2003 or 2004. (Из разг. речи).

б) It looked like the password consisted of *like*, either four or five symbols. (Из разг. речи).

На практике может встретиться случай, когда дискурсный маркер *like* в функции указания на неточность информации используется без конструкции с союзом *or*. В этом случае после дискурсного маркера следует часть предложения, в которой выражена сомнительная информация. Такие предложения часто, но не обязательно завершаются многоточием. Распознавание подобных конструкций облегчается в случаях, когда следующая за *like* конструкция выражена числительным и/или отделяется запятой. Для поиска подобных конструкций используются следующие регулярные выражения, представленные в порядке приоритетности, от наиболее точных к наименее точным:

11) $\backslash \backslash * [like] \{, | \dots\} <[*]> [NUMERAL] * \backslash$

12) $\backslash \backslash * [like] <[*]> [NUMERAL] * \backslash$

13) $\backslash \backslash * [like] \{, | \dots\} * \backslash$

В случае, если предложение соответствует варианту регулярного выражения 13) с многоточием, то на функцию неточности информации может накладываться функция маркера планирования.

Например:

а) I think it was in *like*... 2004. (Из разг. речи).

б) I think it happened *like*, before 1970. (Из разг. речи).

в) I think that the total number of computers in all rooms is *like*, about 150. (Из разг. речи).

В случае, если по регулярным выражениям 1) - 13) невозможно установить дискурсивный статус слова *like*, используется обобщенное регулярное выражение для описания ситуаций, когда слово *like* отделяется с одной или двух сторон запятыми или дефисами, или же когда после слова *like* стоит многоточие, а перед ним – запятая или дефис. В случае, если предложение соответствует такому регулярному выражению, статус слова *like* как дискурсного маркера может быть установлен с высокой степенью точности, поскольку данный дискурсивный маркер имеет свойство сопровождаться паузами в речи, особенно в фокусной функции:

14) $\backslash \backslash * \{, | -\} <[*]> [like] \{, | - | \dots\} * \backslash$

Например:

а) Now I'm going to tell you about, *like*, the most interesting thing that happened to me. (Из разг. речи).

б) It's the most delicious meal that I tasted during, *like*, my whole life! (Из разг. речи).

Как было отмечено ранее, в случаях, когда дискурсивный маркер *like* употреблен в предложении без синтаксических признаков, необходимых и достаточных для его автоматизированного распознавания, возможным является проведение поиска лексических единиц *like* по тексту с указанием ближайшего контекста их употребления и с возможностью быстрого просмотра фрагментов текста, в которых слово *like* было найдено. В данном случае принадлежность слова *like* к классу дискурсивных маркеров и его функция в тексте определяются лингвистом.

Список рекомендуемой литературы

Основная литература

1. Баймуратова, У. Электронный инструментарий переводчика : учебное пособие / У. Баймуратова ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург : ОГУ, 2013. - 120 с. ; То же [Электронный ресурс]. - URL:<http://biblioclub.ru/index.php?page=book&id=259202>
2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

Дополнительная литература

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.
3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>
4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O'Reilly Media, 2012. – 544 p.

Методическая литература

1. Методические рекомендации к самостоятельной работе студентов по дисциплине «Практикум по применению систем автоматизированного анализа текста». Каменский М.В. – Ставрополь, 2026

Интернет-ресурсы

1. Applied Linguistics – <http://www.appliedlinguistics.org>
2. Center for Applied Linguistics – <http://www.cal.org>
3. GATE (Sheffield University, UK) – <http://gate.ac.uk>

Программное обеспечение

1. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>
2. Applied Linguistics – <http://www.appliedlinguistics.org>
3. Center for Applied Linguistics – <http://www.cal.org>
4. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>
5. Emacs – <http://www.gnu.org/software/emacs>
6. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>
7. GNU Compiler Collection – <http://gcc.gnu.org>
8. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>
9. Leopard Language Parser – <http://leopard.loria.fr>
10. NLTK (Natural Language Toolkit) – <http://www.nltk.org>
11. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>
12. Praat – <http://www.fon.hum.uva.nl/praat>
13. Python Programming Language – <http://www.python.org>
14. Sed – <http://www.gnu.org/s/sed>
15. Sonic Visualizer – <http://www.sonicvisualiser.org>
16. Vi Improved - <http://www.vim.org>

Практическое занятие 8.

Тема: Регулярные выражения в Vim (II)

Цель: овладение студентами специфическими практическими навыками и умениями, необходимыми для формализации языковых явлений и построения их формальных моделей, а также использования данных моделей для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Реализуемые компетенции:

ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики
------	--

Актуальность: актуальность изучения темы определяется существованием практической необходимости в подготовке специалистов в области построения формализованных моделей языковых явлений, что обусловлено существенным поворотом в современной лингвистике в сторону использования компьютерных технологий для обработки и анализа текста по формальным признакам.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ СОБЕСЕДОВАНИЯ

Изучите нижеприведенные языковые модели и распишите их с применением метасимволов Vim.

6. Дискурсный маркер *uh/um*.

Слово *uh/um* всегда является дискурсным маркером независимо от занимаемой позиции в предложении, отличаясь только выполняемой функцией, что в значительной степени облегчает его автоматизированное распознавание.

Если дискурсный маркер *uh/um* использован с целью указания на необходимость в дополнительном времени для обдумывания ответа на вопрос, он может быть в большинстве случаев распознан по вопросительному характеру предложения, находящегося до предложения с собственно дискурсным маркером. Таким образом, регулярное выражение для данной функции может быть представлено в следующем виде:

1) `\\ * (? \\ * {uh|um} * \\`

Регулярному выражению 1) будут соответствовать, например, следующие последовательности предложений:

а) A: So, what can you say about it?

B: Well, *uh*, I have to learn more about the situation before I can decide. (Из разг. речи).

б) A: Do you want to go to the theater today?

B: *Uh...* actually, no, I have other plans for the day. (Из разг. речи).

В случае, если предложение, предшествующее предложению, содержащему дискурсный маркер *uh/um*, не носит вопросительного характера, можно сделать вывод о том, что данный маркер указывает на сомнение говорящего или на иную причину, по которой говорящему требуется дополнительное время для обдумывания собственного высказывания, не являющегося ответом на поставленный собеседником вопрос. Регулярное выражение, описывающее указанную ситуацию, отличается от выражения 1) только набором допустимых терминальных знаков препинания в предшествующем предложении:

2) `\\ * (\\.\\.\\.!|) \\ * {uh|um} * \\`

Выражение 2) соответствует, например, следующим наборам предложений:

а) A: Hi!

B: Hey! I, *uh*, wanted to ask you about that CD I gave you yesterday. (Из разг. речи).

б) A: But I'm telling you, his behavior is really irritating!

B: *Uh...* let's talk about something else, this is getting boring. (Из разг. речи).

Поскольку дискурсные маркеры *uh/um* могут иметь различную форму написания - *uh*, *um*, *umm*, *ummm*, *uhh* и т.п., проверка слова на совпадение с шаблоном «`{uh|um}`» организована таким образом, чтобы совпадение возникало и в случае альтернативных форм написания. Таким

же образом учитываются и другие фонетические формы дискурсных маркеров планирования, отличных по звучанию от *uh/um*, например, *err*, *hmm*.

Список рекомендуемой литературы

Основная литература

1. Баймуратова, У. Электронный инструментарий переводчика : учебное пособие / У. Баймуратова ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург : ОГУ, 2013. - 120 с. ; То же [Электронный ресурс]. - URL:<http://biblioclub.ru/index.php?page=book&id=259202>

2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

Дополнительная литература

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.

3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>

4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O’Reilly Media, 2012. – 544 p.

Методическая литература

1. Методические рекомендации к самостоятельной работе студентов по дисциплине «Практикум по применению систем автоматизированного анализа текста». Каменский М.В. – Ставрополь, 2026

Интернет-ресурсы

1. Applied Linguistics – <http://www.appliedlinguistics.org>

2. Center for Applied Linguistics – <http://www.cal.org>

3. GATE (Sheffield University, UK) – <http://gate.ac.uk>

Программное обеспечение

1. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>

2. Applied Linguistics – <http://www.appliedlinguistics.org>

3. Center for Applied Linguistics – <http://www.cal.org>

4. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>

5. Emacs – <http://www.gnu.org/software/emacs>

6. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>

7. GNU Compiler Collection – <http://gcc.gnu.org>

8. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>

9. Leopard Language Parser – <http://leopard.loria.fr>

10. NLTK (Natural Language Toolkit) – <http://www.nltk.org>

11. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>

12. Praat – <http://www.fon.hum.uva.nl/praat>

13. Python Programming Language – <http://www.python.org>

14. Sed – <http://www.gnu.org/s/sed>

15. Sonic Visualizer – <http://www.sonicvisualiser.org>

16. Vi Improved - <http://www.vim.org>

Практическое занятие 9.

Тема: Регулярные выражения в LibreOffice.

Цель: овладение студентами специфическими практическими навыками и умениями, необходимыми для формализации языковых явлений и построения их формальных моделей, а также использования данных моделей для решения научно-исследовательских и научно-практических задач прикладной лингвистики.

Реализуемые компетенции:

ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики
------	--

Актуальность: актуальность изучения темы определяется существованием практической необходимости в подготовке специалистов в области построения формализованных моделей языковых явлений, что обусловлено существенным поворотом в современной лингвистике в сторону использования компьютерных технологий для обработки и анализа текста по формальным признакам.

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ СОБЕСЕДОВАНИЯ

Изучите нижеприведенные языковые модели и распишите их с применением метасимволов LibreOffice.

7. Дискурсный маркер *now*.

В качестве дискурсного маркера слово *now* располагается в начале предложения и выделяется запятой, например:

а) *Now*, in the beginning I said that the code samples provided are relevant for our discussion. (Из разг. речи).

б) *Now*, let's talk about something else, since I'm tired of this conversation. (Из разг. речи).

в) *Now*, I should be very glad if you would have the kindness to show us over the house, Mr. Cunningham (Conan Doyle 1990: 405).

Таким образом, в общем виде синтагматика дискурсного маркера *now* описывается регулярным выражением вида:

1) $\backslash \backslash \langle [^*] \rangle [now] [,] * \backslash \backslash$

Задача поиска данного дискурсного маркера с использованием выражения 1) осложняется тем фактом, что такую же синтагматику имеет слово *now* в определенных случаях, когда оно не является маркером, как, например, в следующем предложении:

Now, though, it will be rather difficult to make everyone understand that Linux isn't any worse than Windows. (Из разг. речи).

Таким образом, выражение 1) указывает только вероятные ситуации употребления *now* как дискурсного маркера. С целью повышения точности поиска дискурсного маркера *now* используются дополнительные регулярные выражения, определяющие контекст, в котором *now* часто выступает в роли дискурсного маркера.

Так, слово *now* с большой вероятностью является дискурсным маркером при следующих условиях.

- Если за ним следует личное местоимение первого лица (I, we). Регулярное выражение, описывающее данную ситуацию, можно представить в виде:

2) $\backslash \backslash \langle [^*] \rangle [now] [,] \{I|we\} * \backslash \backslash$

Например:

а) *Now*, I'd like to speak about the results of your test in the University. (Из разг. речи).

б) *Now*, we can go back to the question we discussed before. (Из разг. речи).

- Если за словом *now* следует конструкция, вводимая словами *as*, *let's*, *let me* и *if* с личным местоимением:

3) $\backslash \backslash \langle [^*] \rangle [now] [,] \{as|if [PERSONAL_PRONOUN]|let\ me|let's\} * \backslash \backslash$

Например:

- а) *Now, as I was saying before, the newest revision of the pixel shaders is present only on the high-end videocards.* (Из разг. речи).
- б) *Now, let me speak about the conflict that happened yesterday.* (Из разг. речи).
- в) *Now, if we go back to what we spoke about before, we'll see that the current situation on the chess board appears to be easily understandable.* (Из разг. Речи).

Список рекомендуемой литературы

Основная литература

1. Баймуратова, У. Электронный инструментарий переводчика : учебное пособие / У. Баймуратова ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург : ОГУ, 2013. - 120 с. ; То же [Электронный ресурс]. - URL:<http://biblioclub.ru/index.php?page=book&id=259202>
2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

Дополнительная литература

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.
3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>
4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O'Reilly Media, 2012. – 544 p.

Методическая литература

1. Методические рекомендации к самостоятельной работе студентов по дисциплине «Практикум по применению систем автоматизированного анализа текста». Каменский М.В. – Ставрополь, 2026

Интернет-ресурсы

1. Applied Linguistics – <http://www.appliedlinguistics.org>
2. Center for Applied Linguistics – <http://www.cal.org>
3. GATE (Sheffield University, UK) – <http://gate.ac.uk>

Программное обеспечение

1. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>
2. Applied Linguistics – <http://www.appliedlinguistics.org>
3. Center for Applied Linguistics – <http://www.cal.org>
4. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>
5. Emacs – <http://www.gnu.org/software/emacs>
6. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>
7. GNU Compiler Collection – <http://gcc.gnu.org>
8. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>
9. Leopard Language Parser – <http://leopard.loria.fr>
10. NLTK (Natural Language Toolkit) – <http://www.nltk.org>
11. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>
12. Praat – <http://www.fon.hum.uva.nl/praat>

13. Python Programming Language – <http://www.python.org>
14. Sed – <http://www.gnu.org/s/sed>
15. Sonic Visualizer – <http://www.sonicvisualiser.org>
16. Vi Improved - <http://www.vim.org>

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Методические указания

по выполнению самостоятельной работы студентов по дисциплине
«Практикум по применению систем автоматизированного анализа текста»

Направление подготовки	<u>45.04.02 Лингвистика</u>
Направленность (профиль)	<u>Современные методы прикладной лингвистики и перевода</u>
Год начала подготовки	<u>2026</u>
Форма обучения	<u>Очная</u>
Реализуется в семестре	<u>2</u>

Ставрополь, 2026

Содержание

1. Введение	3
2. Общая характеристика самостоятельной работы студента при изучении дисциплины.....	4
3. План-график выполнения самостоятельной работы	5
4. Методические рекомендации по изучению теоретического материала	5
5. Методические указания (по видам работ, предусмотренных рабочей программой дисциплины).....	6
6. Список литературы, использованной при составлении методических рекомендаций	11

1. Введение

Методические рекомендации к самостоятельной работе магистрантов по дисциплине «Практикум по применению систем автоматизированного анализа текста» разработаны в соответствии с рабочей программой дисциплины по направлению 45.04.02 - Лингвистика, программа – Современные методы прикладной лингвистики и перевода.

Основной формой работы магистранта является не только работа на лекции, изучение конспекта лекций, их дополнение рекомендованной литературой, но и большая самостоятельная учебная работа, которая позволит глубоко проникнуть в суть рассматриваемой проблемы и подготовить почву для написания кандидатской диссертации. Но для успешной учебной деятельности, ее интенсификации необходимо учитывать следующие субъективные факторы:

1. Знание программного материала, наличие прочной системы знаний, необходимой для усвоения основных дисциплин, предусмотренных программой, общая совокупность которых обуславливает уровень овладения грамматическим компонентом иноязычной речи.

2. Наличие выработанных умений, навыков умственного труда:

а) умение делать глубокий, обстоятельный анализ при работе с книгой, Интернет–источниками;

б) владение логическими операциями: сравнение, анализ, обобщение, определение понятий, правила систематизации и классификации.

3. Специфика познавательных психических процессов: внимание, память, речь, наблюдательность, интеллект и мышление.

4. Хорошая работоспособность, которая обеспечивается нормальным физическим состоянием.

5. Соответствие избранной деятельности, профессии индивидуальным способностям. Необходимо выработать умение саморегулировать свое эмоциональное состояние и устранять обстоятельства, нарушающие деловой настрой, мешающие намеченной работе.

6. Овладение оптимальным стилем работы, обеспечивающим успех в деятельности.

7. Уровень требований к себе, определяемый сложившейся самооценкой.

Адекватная оценка знаний, достоинств, недостатков – важная составляющая самоорганизации человека, без нее невозможна успешная работа по управлению своим поведением, деятельностью.

По наблюдениям исследователей педагогов, одна из основных особенностей обучения заключается в том, что постоянный внешний контроль заменяется самоконтролем, активная роль в обучении принадлежит уже не столько преподавателю, сколько магистранту.

2. Общая характеристика самостоятельной работы студента при изучении дисциплины

Самостоятельная работа магистранта в рамках дисциплины «Практикум по применению систем автоматизированного анализа текста» понимается как планируемая учебная работа, выполняемая во внеаудиторное (аудиторное) время по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Самостоятельная работа направлена на формирование следующих компетенций:

Индекс	Формулировка:
ПК-3	Способен пользоваться профессиональным лингвистическим программным обеспечением и цифровыми технологиями для решения задач в области лингвистики и лингводидактики

Цель самостоятельной работы магистрантов в процессе изучения дисциплины «Практикум по применению систем автоматизированного анализа текста» – научить магистранта осмысленно и самостоятельно работать: 1) с учебным материалом по дисциплине, 2) с научной информацией, актуальными исследованиями в области лингвистики, 3) с эмпирическими данными, получаемыми в ходе экспериментальных лингвистических исследований, 4) с методологическими подходами современных лингвистических исследований; 5) с конкретными лингвистическими методами и методиками.

Задачи самостоятельной работы:

- систематизировать и закрепить полученные теоретические знания и практические умения магистрантов;
- развить познавательные способности и активность магистрантов: творческую инициативу, самостоятельность, ответственность и организованность;
- сформировать и развить навыки ведения самостоятельной работы и овладения методикой исследования при решении разрабатываемых в учебной деятельности проблем и вопросов;
- повысить уровень подготовленности к самостоятельной работе в соответствии с выбранным научным направлением в условиях современного состояния науки и культуры.

Таким образом, самостоятельная работа приобщает научному и исследовательскому творчеству, поиску и анализу актуальных проблем современной психолингвистической науки.

3. План-график выполнения самостоятельной работы.

Код реализуемой компетенции	Вид деятельности студентов	Итоговый продукт самостоятельной работы	Средства и технологии оценки	Объем часов
ПК-3	Изучение литературы	Конспект	Собеседование	2
ПК-3	Подготовка к круглому столу	Индивидуальное задание	Собеседование	2
ПК-3	Работа с электронными ресурсами в сети Интернет	Конспект	Собеседование	2
ПК-3	Подготовка доклада	Доклад и презентация	Собеседование	2
Итого за 2 семестр				8
Итого				8

Для выполнения самостоятельной работы необходимо пользоваться литературой, которая предложена в списке рекомендуемой литературы, Интернет-ресурсами или другими источниками по усмотрению магистранта.

Самостоятельная работа рассчитана на разные уровни мыслительной деятельности. Выполненная работа позволит приобрести не только знания, но и умения, навыки, а также выработать свою методику подготовки, что очень важно в дальнейшем процессе научной деятельности.

При изучении дисциплины предусматриваются следующие формы самостоятельной работы магистранта:

- самостоятельное изучение основной и дополнительной литературы по дисциплине с конспектированием по разделам;
- работа с электронными ресурсами в сети Интернет;
- конспектирование и реферирование первоисточника и научно-исследовательской литературы;
- подготовка к семинару-круглому столу;
- подготовка мультимедийной презентации;
- подготовка доклада.

4. Методические рекомендации по изучению теоретического материала

Чтение основной и дополнительной литературы по курсу с конспектированием по разделам.

Самостоятельная работа при чтении учебной литературы начинается с изучения конспекта материала, полученного при слушании лекций преподавателя. Полученную информацию необходимо осмыслить. При необходимости, в конспект лекций могут быть внесены схемы, другая дополнительная информация. При изучении нового материала составляется конспект. Сжато излагается самое существенное в данном материале.

Работа с электронными ресурсами в сети Интернет.

Для повышения эффективности самостоятельной работы магистрант должен уметь работать в поисковой системе сети Интернет и использовать найденную информацию при подготовке к занятиям. Поиск информации можно вести по автору, заглавию, виду издания, году издания или издательству. Также в сети Интернет доступна услуга по скачиванию методических указаний и учебных пособий, подбору необходимой научной литературы.

Конспектирование и реферирование первоисточника и научно-исследовательской литературы.

Конспект представляет собой дословные выписки из текста источника. При этом необходимо понимать, что конспект – это не полное переписывание чужого текста. Необходимо знать, что при написании конспекта сначала прочитывается текст – источник, в нём выделяются основные положения, подбираются примеры, идёт перекомпоновка материала, а уже затем оформляется текст конспекта. Конспект может быть полным, когда работа идёт со всем текстом источника или неполным, когда интерес представляет какой-либо один или несколько вопросов, затронутых в источнике.

Реферирование — это сложный творческий процесс, в основе которого лежит умение выделить главную информацию из текста первоисточника. Реферирование – процесс аналитически-синтетической обработки информации, которая заключается в анализе первичного документа, нахождении значимых в смысловом отношении данных (основных положений, фактов, доведите день, результатов, выводов) Реферирование имеет целью сократить физический объем первичного документа при сохранении его основного смыслового содержания, используется в научной, издательской, информационной и библиографической деятельности.

5. Методические указания (по видам работ, предусмотренных рабочей программой дисциплины)

Подготовка к круглому столу

Подготовка к семинару-круглому столу начинается с распределение форм участия и функции магистрантов в семинаре-круглом столе. Магистрантами осуществляется определение круга проблем и вопросов, подлежащих обсуждению; подбор основной и дополнительной литературы к теме семинара - круглого стола, а также дальнейшее изучение литературы.

Подготовка мультимедийной презентации

Презентация, согласно толковому словарю русского языка Д.Н. Ушакова: «... способ подачи информации, в котором присутствуют рисунки, фотографии, анимация и звук». Для подготовки презентации рекомендуется

использовать LibreOffice Impress (для подготовки собственно мультимедийных презентаций) и LibreOffice Writer (для составления текстового сопровождения презентации), являющихся компонентами открытого и свободного офисного пакета LibreOffice. Также допускается использование проприетарного продукта Microsoft Office (Powerpoint и Word, соответственно), однако в этом случае должны использоваться наиболее совместимые форматы .ppt, .doc (но не .pptx, .docx).

Для подготовки презентации необходимо собрать и обработать начальную информацию.

Последовательность подготовки презентации:

1. Четко сформулировать цель презентации: вы хотите свою аудиторию мотивировать, убедить, заразить какой-то идеей или просто формально отчитаться.
2. Определить каков будет формат презентации: живое выступление (тогда, сколько будет его продолжительность) или электронная рассылка (каков будет контекст презентации).
3. Отобрать всю содержательную часть для презентации и выстроить логическую цепочку представления.
4. Определить ключевые моменты в содержании текста и выделить их.
5. Определить виды визуализации (картинки) для отображения их на слайдах в соответствии с логикой, целью и спецификой материала.
6. Подобрать дизайн и форматировать слайды (количество картинок и текста, их расположение, цвет и размер).
7. Проверить визуальное восприятие презентации.

К видам визуализации относятся иллюстрации, образы, диаграммы, таблицы.

Иллюстрация – представление реально существующего зрительного ряда.

Образы – в отличие от иллюстраций – метафора. Их назначение – вызвать эмоцию и создать отношение к ней, воздействовать на аудиторию. С помощью хорошо продуманных и представляемых образов, информация может надолго остаться в памяти человека.

Диаграмма – визуализация количественных и качественных связей. Их используют для убедительной демонстрации данных, для пространственного мышления в дополнение к логическому.

Таблица – конкретный, наглядный и точный показ данных. Ее основное назначение – структурировать информацию, что порой облегчает восприятие данных аудиторией.

Практические советы по подготовке презентации.

- готовьте отдельно: печатный текст + слайды + раздаточный материал;
- слайды – визуальная подача информации, которая должна содержать
- минимум текста, максимум изображений, несущих смысловую

- нагрузку, выглядеть наглядно и просто;
- текстовое содержание презентации – устная речь или чтение, которая
- должна включать аргументы, факты, доказательства и эмоции;
- рекомендуемое число слайдов 10-12;
- обязательная информация для презентации: тема, фамилия и инициалы
- выступающего; план сообщения; краткие выводы из всего сказанного; список использованных источников;
- раздаточный материал – должен обеспечивать ту же глубину и охват, что и живое выступление: люди больше доверяют тому, что они могут унести с собой, чем исчезающим изображениям, слова и слайды забываются, а раздаточный материал остается постоянным осязаемым напоминанием; раздаточный материал важно раздавать в конце презентации; раздаточный материалы должны отличаться от слайдов, должны быть более информативными.

Доклад, согласно толковому словарю русского языка Д.Н. Ушакова:

«... сообщение по заданной теме, с целью внести знания из дополнительной литературы, систематизировать материал, проиллюстрировать примерами, развивать навыки самостоятельной работы с научной литературой, познавательный интерес к научному познанию».

Тема доклада должна быть согласована с преподавателем и соответствовать теме учебного занятия. Материалы при его подготовке, должны соответствовать научно-методическим требованиям вуза и быть указаны в докладе. Необходимо соблюдать регламент, оговоренный при получении задания. Иллюстрации должны быть достаточными, но не чрезмерными.

Работа магистранта над докладом-презентацией включает отработку умения самостоятельно обобщать материал и делать выводы в заключении, умения ориентироваться в материале и отвечать на дополнительные вопросы слушателей, отработку навыков ораторства, умения проводить диспут.

Докладчики должны знать и уметь: сообщать новую информацию; использовать технические средства; хорошо ориентироваться в теме всего семинарского занятия; дискутировать и быстро отвечать на заданные вопросы; четко выполнять установленный регламент (не более 10 минут); иметь представление о композиционной структуре доклада и др.

Структура выступления

Вступление помогает обеспечить успех выступления по любой тематике. Вступление должно содержать: название, сообщение основной идеи, современную оценку предмета изложения, краткое перечисление рассматриваемых вопросов, живую интересную форму изложения, акцентирование внимания на важных моментах, оригинальность подхода.

Основная часть, в которой выступающий должен глубоко раскрыть суть затронутой темы, обычно строится по принципу отчета. Задача основной части – представить достаточно данных для того, чтобы слушатели заинтересовались темой и захотели ознакомиться с материалами. При этом логическая структура теоретического блока не должны даваться без наглядных пособий, аудиовизуальных и визуальных материалов.

Заключение – ясное, четкое обобщение и краткие выводы, которых всегда ждут слушатели

Написание доклада

Доклад – публичное сообщение, представляющее собой развёрнутое изложение определённой темы.

Этапы подготовки доклада:

1. Определение цели доклада.
2. Подбор необходимого материала, определяющего содержание доклада.
3. Составление плана доклада, распределение собранного материала в необходимой логической последовательности.
4. Общее знакомство с литературой и выделение среди источников главного.
5. Уточнение плана, отбор материала к каждому пункту плана.
6. Композиционное оформление доклада.
7. Заучивание, запоминание текста доклада, подготовки тезисов выступления.
8. Выступление с докладом.
9. Обсуждение доклада.
10. Оценивание доклада

Композиционное оформление доклада – это его реальная речевая внешняя структура, в ней отражается соотношение частей выступления по их цели, стилистическим особенностям, по объёму, сочетанию рациональных и эмоциональных моментов, как правило, элементами композиции доклада являются: вступление, определение предмета выступления, изложение (опровержение), заключение.

Вступление помогает обеспечить успех выступления по любой тематике.

Вступление должно содержать:

- название доклада;
- сообщение основной идеи;
- современную оценку предмета изложения;
- краткое перечисление рассматриваемых вопросов;
- интересную для слушателей форму изложения;
- акцентирование оригинальности подхода.

Выступление состоит из следующих частей:

Основная часть, в которой выступающий должен раскрыть суть темы,

обычно строится по принципу отчёта. Задача основной части: представить достаточно данных для того, чтобы слушатели заинтересовались темой и захотели ознакомиться с материалами.

Заключение - это чёткое обобщение и краткие выводы по излагаемой теме.

Тематика докладов

- 1) Основные направления прикладной лингвистики. Методология исследования.
- 2) Автоматизированная обработка неструктурированной текстовой информации на естественном языке.
- 3) Распознавание аудио- и видеоинформации.
- 4) Текстовые редакторы с интегрированными средствами массовой обработки текстовой информации на естественных языках: vi, emacs.
- 5) Искусственный интеллект в автоматизированной обработке текста. Четкая и нечеткая логика. Нейронные сети.
- 6) Регулярные выражения как методологическое основание массовой обработки текстовых массивов на естественных языках.
- 7) Автоматизированный лексический и синтаксический анализ текста.
- 8) Свободное программное обеспечение с открытым кодом в прикладной лингвистике.
- 9) Системы автоматизированного перевода текста. Современное состояние проблемы. Преимущества и недостатки технологии.

Тематика обсуждений в форме «круглого стола»

- 1) Проблемы искусственного интеллекта: достоинства и недостатки современных подходов к созданию ИИ, ограничения, проблемы и их решения.
- 2) Компьютерная лингвистика как квантитативный подход к анализу текстовых данных в лингвистике.
- 3) Ресурсы и программное обеспечение для обработки текстовых документов на естественных языках.
- 4) Пользовательский интерфейс в современных средах автоматизированного анализа текста. Пути модернизации, проблемы и их решения.
- 5) Предиктивные методы редактирования и анализа текста.
- 6) Современное состояние проблемы автоматизированного анализа текста на естественных языках.
- 7) Машинный перевод: аргументы «за» и «против», проблемы и возможные пути их решения.
- 8) Свободное программное обеспечение vs. проприетарное программное обеспечения: достоинства и недостатки профессионального использования в сфере прикладной лингвистики.

6. Список литературы, использованной при составлении методических рекомендаций

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. Каменский М.В. Когнитивно-функциональная модель дискурсных маркеров. – Ставрополь: Изд-во СКФУ, 2014. – 186 с.
3. 5th GATE Training Course. - The University of Sheffield, 2012. - <http://gate.ac.uk/wiki/TrainingCourseJune2012/>
4. Friedl, J.E.F. Mastering Regular Expressions / J.E.F. Friedl. – 3rd. ed. O'Reilly Media, 2012. – 544 p.
5. Роббинс, А. Изучаем редакторы vi и Vim / А. Роббинс, Э. Хана, Л. Лэмб. – М.: Символ-Плюс, 2013. – 512 с.
6. Horstmann, C., Cornell, G. Core Java Volume I – Fundamentals (9th Edition) / C. Horstmann, G. Cornell. – Prentice Hall, 2012. – 1008 p.
7. Annotation Graph Toolkit (AGTK) – <http://agtk.sourceforge.net>
8. Applied Linguistics – <http://www.appliedlinguistics.org>
9. Center for Applied Linguistics – <http://www.cal.org>
10. Code::Blocks C/C++ IDE – <http://www.codeblocks.org>
11. Emacs – <http://www.gnu.org/software/emacs>
12. GATE (General Architecture for Text Engineering) – <http://gate.ac.uk>
13. GNU Compiler Collection – <http://gcc.gnu.org>
14. Helsinki Finite-State Technology – <http://sourceforge.net/projects/hfst>
15. Leopard Language Parser – <http://leopard.loria.fr>
16. NLTK (Natural Language Toolkit) – <http://www.nltk.org>
17. Open Natural Language Processing (OpenNLP) – <http://opennlp.sourceforge.net>
18. Praat – <http://www.fon.hum.uva.nl/praat>
19. Python Programming Language – <http://www.python.org>
20. Sed – <http://www.gnu.org/s/sed>
21. Sonic Visualizer – <http://www.sonicvisualiser.org>
22. Vi Improved - <http://www.vim.org>