

МИНОБРНАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНСТИТУТ ЧЕЛОВЕКА ЦИФРОВОЙ ЭПОХИ

УТВЕРЖДАЮ

Руководитель САЕ _____ Г.В. Можаява

" ____ " _____ 2018 г.

Рабочая программа дисциплины

ВВЕДЕНИЕ В АНАЛИЗ ЕСТЕСТВЕННОГО ЯЗЫКА (NLP)¹

Направление подготовки

45.04.03 Фундаментальная и прикладная лингвистика

Профиль подготовки

Компьютерная и когнитивная лингвистика

Квалификация (степень) выпускника

Магистр

Форма обучения

очная

**БЛАГОТВОРИТЕЛЬНЫЙ
ФОНД В. ПОТАНИНА**



Томск-2018

¹ Программа создана при поддержке Благотворительного Фонда В. Потанина

1. Код и наименование дисциплины (модуля)

В.1.3.2 Введение в анализ естественного языка (NLP)

2. Место дисциплины в структуре ООП магистратуры

Является обязательной дисциплиной вариативной части В.1.3 профиля «Компьютерная и когнитивная лингвистика».

3. 1 год обучения, I семестр**4. Входные требования для освоения модуля дисциплины, предварительные условия**

Базовыми для данной дисциплины для студентов I курса являются следующие курсы или их эквиваленты:

- Иностранный язык (английский);
- Введение в теорию языка (или аналогичный);
- Общее языкознание (или аналогичный);

Основными требованиями к «входным» знаниям, умениям и готовностям обучающегося и сформированным компетенциям, необходимым при освоении дисциплины «Введение в анализ естественного языка (NLP)» для студентов I курса, приобретенным в результате освоения предшествующих дисциплин являются следующие.

Студент должен

знать:

основные теоретические положения современного языкознания; основные факты об устройстве языка; базовые понятия фонетики и фонологии, морфологии, синтаксиса, дискурса-анализа; базовую терминологию языкознания;

уметь:

работать в коллективе, толерантно воспринимать социальные, этнические, конфессиональные и культурные различия;
корректно вести научную дискуссию по заданной проблеме;
составлять реферативные обзоры, сообщения, аннотированные библиографии на заданную тему;

владеть:

навыками восприятия устных и письменных текстов научного стиля на родном языке;
навыками восприятия письменных текстов научного стиля в пределах темы курса на английском языке;
навыками лингвистического анализа на различных уровнях языковой системы.

5. Общая трудоемкость модуля дисциплины составляет 3 зачетных единицы, 108 часов, из которых 26 часов составляет контактная работа обучающихся с преподавателем (лекционные занятия, групповые практические занятия с преподавателем, индивидуальные консультации). 46 часов составляет самостоятельная работа обучающихся. 36 часов отводятся на подготовку к экзамену.

6. Формат обучения: аудиторные занятия.**7. Планируемые результаты обучения по модулю дисциплины, соотнесенные с планируемыми результатами освоения образовательной деятельности**

Формируемые компетенции (код компетенции, уровень (этап) освоения)	Планируемые результаты обучения по дисциплине (модулю)
ОПК-4, I уровень	<i>Знать:</i> Современное состояние в области компьютерной лингвистики и информационных технологий, основные направления и тенденции в данной сфере, ключевые проблемы, которые решаются на современном этапе развития

	<p>лингвистических информационных технологий и подходы к их решению</p> <p><i>Уметь:</i> работать с лингвистическими программными продуктами для решения прикладных и исследовательских задач</p> <p><i>Владеть:</i> навыками применения лингвистических программных продуктов, оценки уместности их использования в конкретной ситуации и ограничений, связанных с их использованием</p>
ПК-1, I уровень	<p><i>Знать:</i> основные теоретические и прикладные направления науки в области теории языка, конкретных языков, прикладной и компьютерной лингвистики, основные понятия соответствующего научного направления, методологические основания планируемого исследования</p> <p><i>Уметь:</i> применять теоретические знания при создании собственного исследования, в том числе при лингвистическом анализе материала; применять методологию соответствующего направления лингвистики в собственных научных исследованиях; работать с научной литературой в области теории языка, конкретных языков, прикладной и компьютерной лингвистики</p> <p><i>Владеть:</i> основными методами и приемами исследовательской работы как одного из видов профессиональной деятельности, в том числе методами лингвистического анализа в области теории языка, лингвистики конкретных языков, прикладной и компьютерной лингвистики; навыками работы с разными типами источников</p>
ПК-7, I уровень	<p><i>Знать:</i> Общие принципы структуры и функционирования лингвистических баз данных, устройства печатных и электронных словарей, базовые параметры корпусов, базовую логику построения онтологий; области применения данных продуктов</p> <p><i>Уметь:</i> заполнять базы данных, программировать словарные статьи; заполнять билингвальный корпус;</p> <p><i>Владеть:</i> навыком заполнения словарей, корпуса родного языка с использованием интерфейса; навыком правки существующей онторазметки;</p>
ПК-8, I уровень	<p><i>Знать:</i> общие принципы структуры и функционирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, устройства лингвистических компоненты интеллектуальных и информационных электронных систем, их базовые параметры, логику и области применения данных продуктов</p> <p><i>Уметь:</i> написать заявку в техническую поддержку интеллектуальной информационной системы; написать предложение по улучшению и расширению функционала существующей системы с лингвистическим обоснованием</p> <p><i>Владеть:</i> навыком извлечения знаний из интеллектуальной информационной системы; распознавания речи с помощью доступных программных платформ; написания отчета по сравнению результатов работы с различными системами</p>

8. Содержание дисциплины (модуля) и структура учебных видов деятельности

Наименование разделов и тем	Всего (час.)	Контактная работа (час.)		Самостоятельная работа (час.)
		Лекция	Семинар	
1. Введение в NLP (I): теоретические проблемы, соотношение NLP и лингвистики, основные источники	2	2		
2. Введение в NLP (II): N-граммы и векторная семантика, корпусные технологии	2	2		
3. Введение в NLP (III): регулярные выражения, нормализация текста,	2	2		

расстояние редактирования. Частеречная разметка.				
4. Математические методы (I): конечные автоматы и трансдукторы, Naive Bayes, k-nearest neighbor, artificial neural networks			2	4
5. Математические методы (II): Математические методы (II): decision trees, random forest, hidden Markov models			2	4
6. Математические методы (III): support vector machines, Principal Component Analysis			2	4
7. Сентимент-анализ (I): современное состояние поля, реализации Bag-of- Words и более продвинутых подходов			2	4
8. Сентимент-анализ (II): Решения в области, проприетарные и открытые; продукты для мониторинга соцсетей; использование метаданных			2	4
9. Разрешение анафоры в тексте: лингвистический и вычислительный аспект			2	4
10. Машинный перевод: rule-based, NMT: seq2seq и другие модели			2	4
11. Speech Recognition (I): фонетическая структура языка, акустика речи, восприятие и производство речи			2	6
12. Speech Recognition (II): машинное обучение и распознавание речи			2	6
13. Нормализация и подготовка текста к обработке: text pre-processing, POS- tagging, parsing			2	6
Экзамен	36			
Итого	108	6	20	46

9. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по модулю и методические указания для обучающихся по освоению модуля дисциплины.

1. Колесникова Н.И. От конспекта к диссертации: Учебное пособие по развитию навыков письменной речи. — М.: Флинта: Наука, 2011. — 287 с.
2. Комарова З. И. Методология, метод, методика и технология научных исследований в лингвистике: учебное пособие. – Екатеринбург: Изд-во УрФУ, 2012. – 817 с.
3. Крылова В.С., Григорьевская С.М., Кичигина Е.Ю. Рекомендации. Библиографическое описание документа. Общие требования и правила составления. – Электронный ресурс: URL: <http://www.lib.tsu.ru/win/produkcija/metodichka/metodich.html>.
4. Эко У. Как написать дипломную работу. – М. Университет, 2003. – 238 с.
5. Steen, G. A method for linguistic metaphor identification: from MIP to MIPVU [Text] / G.J.

Steen et al. – John Benjamins B.V., 2010. – 253 p.

В приведенной ниже таблице и приложении к ней представлены материалы для обеспечения самостоятельной работы студентов и методические указания по ее подготовке.

Тема	Неделя семестра	Задания для самостоятельной работы
14. Математические методы (I): конечные автоматы и трандукторы, Naive Bayes, k-nearest neighbor, artificial neural networks	4	<p>Студенты готовят презентацию по одному из указанных методов, руководствуясь следующей структурой:</p> <ol style="list-style-type: none"> 1. Небольшая история метода, в т.ч. с изложением его математических и когнитивных — для нейронных сетей — основ в весьма упрощенном виде. 2. Компоненты метода. 3. Процесс и особенности обучения. 4. Варианты метода. 5. Сферы применения 6. Критика/недостатки/особенности, которые нужно иметь в виду 7. Существующие решения: готовые к применению и открытые 8. Применение в NLP, м.б. какой-то интересный проект/"кейс"/статья вкратце.
15. Математические методы (II): Математические методы (II): decision trees, random forest, hidden Markov models	5	<p>Студенты готовят презентацию по одному из указанных методов, руководствуясь следующей структурой:</p> <ol style="list-style-type: none"> 1. Небольшая история метода, в т.ч. с изложением его математических и когнитивных — для нейронных сетей — основ в весьма упрощенном виде. 2. Компоненты метода. 3. Процесс и особенности обучения. 4. Варианты метода. 5. Сферы применения 6. Критика/недостатки/особенности, которые нужно иметь в виду 7. Существующие решения: готовые к применению и открытые 8. Применение в NLP, м.б. какой-то интересный проект/"кейс"/статья вкратце.

<p>16. Математические методы (III): support vector machines, Principal Component Analysis</p>	6	<p>Студенты готовят презентацию по одному из указанных методов, руководствуясь следующей структурой:</p> <ol style="list-style-type: none"> 1. Небольшая история метода, в т.ч. с изложением его математических и когнитивных — для нейронных сетей — основ в весьма упрощенном виде. 2. Компоненты метода. 3. Процесс и особенности обучения. 4. Варианты метода. 5. Сферы применения 6. Критика/недостатки/особенности, которые нужно иметь в виду 7. Существующие решения: готовые к применению и открытые 8. Применение в NLP, м.б. какой-то интересный проект/"кейс"/статья вкратце.
<p>17. Сентимент-анализ (I): современное состояние поля, реализации Bag-of-Words и более продвинутых подходов</p>	7	<p>Подготовка к семинару. Цели, методы и материалы сентимент-анализа.</p> <p>Sentiment Analysis and Subjectivity // Handbook of Natural Language Processing. / Eds. Nitin Indurkha, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.</p> <p>Taboada M. et al. Lexicon-based methods for sentiment analysis // Computational linguistics. – 2011. – Т. 37. – №. 2. – С. 267-307.</p> <p>Benamara F., Taboada M., Mathieu Y. Evaluative language beyond bags of words: Linguistic insights and computational applications // Computational linguistics. – 2017. – Т. 43. – №. 1. – С. 201-264.</p> <p>Краткое эссе (ок. 3000 знаков): <i>5 областей применения сентимент-анализа.</i></p>
<p>18. Сентимент-анализ (II): Решения в области, проприетарные и открытые; продукты для мониторинга соцсетей; использование метаданных</p>	8	<p>Подготовка к семинару. Цели, методы и материалы сентимент-анализа.</p> <p>Sentiment Analysis and Subjectivity // Handbook of Natural Language Processing. / Eds. Nitin Indurkha, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.</p> <p>Taboada M. et al. Lexicon-based methods for sentiment analysis // Computational linguistics. – 2011. – Т. 37. – №. 2. – С. 267-307.</p> <p>Benamara F., Taboada M., Mathieu Y. Evaluative language beyond bags of words: Linguistic insights and computational applications // Computational linguistics. – 2017. – Т. 43. – №. 1. – С. 201-264.</p> <p>Краткое эссе (ок. 3000 знаков): <i>5 областей применения сентимент-анализа.</i></p>

<p>19. Разрешение анафоры в тексте: лингвистический и вычислительный аспект</p>	9	<p>Подготовка к семинару. Конспектирование и сравнение основных типов формальных грамматик, подходов к выделению частей речи.</p> <p>Syntactic Parsing // Handbook of Natural Language Processing. / Eds. Nitin Indurkha, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.</p> <p>Treebank Annotation // Handbook of Natural Language Processing. / Eds. Nitin Indurkha, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.</p> <p>Part-of-Speech Tagging // Handbook of Natural Language Processing. / Eds. Nitin Indurkha, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.</p> <p>Statistical Parsing // Handbook of Natural Language Processing. / Eds. Nitin Indurkha, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.</p> <p>Segmentation and Morphology // The handbook of computational linguistics and natural language processing / Eds Alexander Clark, Chris Fox, Shalom Lappin. – Wiley-Blackwell, 2010. – 801 p.</p> <p>Formal Grammars of English // Jurafsky D., Martin J.H. Speech and Language Processing (3rd ed. draft). – URL: https://web.stanford.edu/~jurafsky/slp3/11.pdf</p> <p>Syntactic Parsing // Jurafsky D., Martin J.H. Speech and Language Processing (3rd ed. draft). – URL: https://web.stanford.edu/~jurafsky/slp3/12.pdf</p> <p>Поиск существующих решений для русского языка.</p>
<p>20. Машинный перевод: rule-based, NMT: seq2seq и другие модели</p>	10	<p>Подготовка к семинару. Machine Translation // The handbook of computational linguistics and natural language processing / Eds Alexander Clark, Chris Fox, Shalom Lappin. – Wiley-Blackwell, 2010. – 801 p.</p> <p>Lembersky G., Ordan N., Wintner S. Language models for machine translation: Original vs. translated texts // <i>Computational Linguistics</i>. – 2012. – Т. 38. – №. 4. – С. 799-825.</p> <p>Письменный обзор существующих программных продуктов в области машинного перевода.</p>
<p>21. Speech Recognition (I): фонетическая структура языка, акустика речи, восприятие и производство речи</p>	11	<p>Подготовка двадцатиминутного сообщения по выбранному заранее методу с использованием источников, указанных в списках основной и дополнительной литературы.</p>
<p>22. Speech Recognition (II): машинное обучение и распознавание речи</p>	12	<p>Подготовка двадцатиминутного сообщения по выбранному заранее методу с использованием источников, указанных в списках основной и дополнительной литературы.</p>

23. Нормализация и подготовка текста к обработке: text pre-processing, POS-tagging, parsing	13	Подготовка двадцатиминутного сообщения по выбранному заранее методу с использованием источников, указанных в списках основной и дополнительной литературы.
---	----	--

Приложение к таблице «Самостоятельная работа студентов»

Предлагаемая структура сообщения и презентации о формальном методе анализа естественного языка (темы 11-13):

1. История метода (время возникновения, связь с математическими концепциями и областями, авторы);
2. Математическая сущность метода;
3. Ограничения на применимость метода (материал, специфика решаемых задач);
4. Практические применения метода (набор задач в области NLP, в решении которых в настоящий момент применяется данный метод), конкретные примеры;
5. Эффективность метода.

Математическая сторона метода может быть сокращена до необходимого для базового понимания объема. Фокус презентации рекомендуется смещать на практическое применение в подзадачах обработки естественного языка; желательно углубленно продемонстрировать хотя бы одну реальную реализацию метода и полученные результаты (на материале существующих исследований).

Устная презентация должна отвечать следующим критериям:

1. успешное удержание внимания на речи и презентации докладчика;
2. адекватное оформление презентационных материалов по времени, дизайну и структуре;
3. разделение поданной информации на главную и второстепенную;
4. умение адаптировать презентацию к нуждам аудитории;
5. умение корректно (в рамках научной дискуссии) отвечать на поставленные вопросы аудитории и сделанные замечания.

10. Форма промежуточной аттестации и фонд оценочных средств:

- **Перечень компетенций выпускников образовательной программы, в формировании которых участвует дисциплина и их «карты»**

способность к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);

способность проводить самостоятельные исследования и получать новые научные результаты в области теории языка, лингвистики конкретных языков, прикладной и компьютерной лингвистики (ПК-1);

способность разрабатывать лингвистические компоненты электронных языковых ресурсов (лингвистические корпуса, словари, онтологии, базы данных) (ПК-7);

способность разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты интеллектуальных и информационных электронных систем (ПК-8).

- **Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения, характеризующих этапы формирования компетенций**

Анализ и оценивание существующих исследований на предмет охвата материала, качества анализа и теоретических оснований. Представление критического анализа в устном выступлении на семинаре. (ОПК-4)-I (знания), (ПК-1)-I (знания), (ПК-7)-I (знания), (ПК-8)-I (знания)

Критерии оценивания: работа может быть зачтена, если анализ охватывает не менее 5 работ по заданной теме, выделяет и сравнивает методологию исследований, проводит теоретическое обобщение работ, отмечает как их слабые, так и сильные стороны. На устное представление анализа отводится 5-7 минут.

Написание эссе на темы, предложенные в приложении в разделе «Самостоятельная работа» или иным (по предложению студента/преподавателя). (ОПК-4)-I (знания), (ПК-1)-I (знания), (ПК-7)-I (знания), (ПК-8)-I (знания)

Критерии оценивания: работа может быть зачтена, если соответствует критериям жанра эссе и показывает как широкое знакомство с работами по выбранной проблематике, так и сложившуюся позицию по выбранному вопросу. Позиция должна быть аргументирована не менее, чем тремя доводами; приветствуется предвосхищение позиции оппонента.

Планирование исследования: студенты получают практическое задание в области компьютерной лингвистики, например, «Чат-бот «Психолог» для студентов университета» и должны разработать план выполнения с конкретными источниками литературы, аналогичными проектами и шагами реализации. (ОПК-4)-I (знания), (ПК-1)-I (знания), (ПК-7)-I (знания), (ПК-8)-I (знания)

Критерии оценивания: работа может быть зачтена, если после коррекции и обсуждения с преподавателем построен четко реализуемый план работы, определены источники и проекты, относительно которых будет оцениваться функционирование потенциального программного продукта.

Подготовка аннотированной библиографии (10-15 источников) к одной из тем семинаров. (ОПК-4)-I (знания), (ПК-1)-I (знания), (ПК-7)-I (знания), (ПК-8)-I (знания)

Критерии оценивания: работа может быть зачтена, если аннотация верно фиксирует суть источника, отмечает наиболее существенные особенности; аннотации не обязательно должны следовать единому шаблону.

Комбинированный индивидуальный/парный/групповой семинар: студентам предъявляется тема для мозгового штурма, например, «Какое приложение может иметь теория функциональной грамматики в области семантических репрезентаций?». Дается пять минут на индивидуальную подготовку, затем студенты объединяются в пары и выбирают три лучших идеи (следующие пять минут). После этого группа выбирает 3 лучших идеи и представляет их преподавателю. (ОПК-4)-I (знания), (ПК-1)-I (знания), (ПК-7)-I (знания), (ПК-8)-I (знания)

Критерии оценивания: работа может быть зачтена, если высказанные гипотезы представляют интерес для современного состояния науки о языке и мышлении и являются проверяемыми.

Задания на определение ключевых понятий различных областей NLP. (ОПК-4)-I (знания), (ПК-1)-I (знания), (ПК-7)-I (знания), (ПК-8)-I (знания)

Критерии оценивания: работа может быть зачтена, если выделены основные ключевые понятия концепции и установлены связи между ними, даны полноценные определения; приветствуется нахождение иноязычных терминологических аналогов, проведение параллелей с понятиями

других известных студенту концепций.

Лабораторная работа по заполнению словаря, программированию словарных статей, корпусов различных типов, правке онторазметки, извлечению знаний из интеллектуальной информационной системы, распознаванию речи с помощью доступных программных платформ.

Критерии оценивания: работа может быть зачтена, если задание выполнено в объеме не менее 80% от установленного преподавателем, не содержит принципиальных ошибок, мелкие ошибки устранены. (ОПК-4)-I (владение), (ПК-1)-I (владение), (ПК-7)-I (владение), (ПК-8)-I (владение), (ОПК-4)-I (умения), (ПК-1)-I (умения), (ПК-7)-I (умения), (ПК-8)-I (умения)

Написанию отчета по сравнению результатов работы с различными системами, написанию заявки в техническую поддержку интеллектуальной информационной системы; написание предложений по улучшению и расширению функционала существующих систем. (ОПК-4)-I (владение), (ПК-1)-I (владение), (ПК-7)-I (владение), (ПК-8)-I (владение), (ОПК-4)-I (умения), (ПК-1)-I (умения), (ПК-7)-I (умения), (ПК-8)-I (умения)

Критерии оценивания: работа может быть зачтена, если отчет или предложения демонстрируют освоенность и уверенный уровень использования систем, критический подход к их функциям и интерфейсу.

Методические материалы, определяющие процедуры оценивания результатов обучения

Экзамен состоит в выполнении практического задания по позиционированию своего магистерского исследования в рамках поля NLP. Работа выполняется письменно и оформляется согласно требованиям, предъявляемым к курсовым работам. Объем работы – до 15 стр. А4. Защита работы происходит на экзамене в форме презентации и ответа на вопросы экзаменатора и присутствующих.

Работа и презентация может быть оценена, если она соответствует следующим критериям.

Работа демонстрирует наличие у студента знаний о поле NLP, основных и специальных методах и приемах анализа естественного языка. Магистрант должен проявить способность позиционировать свою работу в контексте современной практики обработки естественного языка. Работа должна излагать суть планируемого проекта магистранта и его место в современной парадигме NLP. Приветствуется наличие детализированного плана разработки в соотнесении с необходимыми для его реализации технологиями.

Работа должна отвечать специфике научного стиля, быть ясно изложена. Задание, выполненное студентом, должно демонстрировать знание принципов создания и презентации научного текста в форме плана проекта, реферата, плана исследования, научного эссе, их структурирования и оформления. Особое внимание уделяется корректности оформления текста, иллюстраций и ссылок по существующему ГОСТу.

Требования к презентации:

1. успешное удержание внимания на речи и презентации докладчика;
2. адекватное оформление презентационных материалов по времени, дизайну и структуре;
3. разделение поданной информации на главную и второстепенную;
4. умение адаптировать презентацию к нуждам аудитории;
5. умение корректно (в рамках научной дискуссии) отвечать на поставленные вопросы

аудитории и сделанные замечания.

Оценка «удовлетворительно» может быть поставлена при выполнении 60 процентов самостоятельной работы в семестре, выполнении проектной работы, фрагментарно описывающей магистерскую проектную работу, отражающей недостаточное знание поля NLP, технологий обработки языка и современного состояния субдисциплин. Презентация работы выполнена с нарушением критериев, ответы на вопросы даны некорректно или не по существу.

Оценка «хорошо» может быть поставлена при выполнении 80 процентов самостоятельной работы в семестре, выполнении проектной работы, в целом описывающей магистерскую проектную работу, отражающей знание поля NLP, основных технологий обработки языка и современного состояния субдисциплин. Презентация работы выполнена согласно критериям, возможны недочеты в изложении содержания работы, затруднения при ответах на вопросы.

Оценка «отлично» может быть поставлена при выполнении 95-100 процентов самостоятельной работы в семестре, выполнении проектной работы, исчерпывающе описывающей магистерскую проектную работу, отражающей полноценное знание поля NLP, технологий обработки языка и современного состояния субдисциплин. Презентация полностью отвечает всем критериям, ответы на вопросы экзаменатора и присутствующих демонстрируют знание специфики релевантной области NLP.

Комментарий: следующие методические пособия дают критерии оценки правильности формулирования квалификационных параметров, оформления библиографических ссылок и текста студенческой работы.

1. Колесникова Н.И. От конспекта к диссертации: Учебное пособие по развитию навыков письменной речи. — М.: Флинта: Наука, 2011. — 287 с.
2. Крылова В.С., Григорьевская С.М., Кичигина Е.Ю. Рекомендации. Библиографическое описание документа. Общие требования и правила составления. — Электронный ресурс: URL: <http://www.lib.tsu.ru/win/produkcija/metodichka/metodich.html>.

11. Ресурсное обеспечение:

Основная учебная литература:

1. Handbook of Natural Language Processing. / Eds. Nitin Indurkha, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.
2. Jurafsky D., Martin J.H. Speech and Language Processing (3rd ed. draft). – URL: <https://web.stanford.edu/~jurafsky/slp3/>
3. The handbook of computational linguistics and natural language processing / Eds Alexander Clark, Chris Fox, Shalom Lappin. – Wiley-Blackwell, 2010. – 801 p.

Научные журналы:

1. *ACL Anthology*. URL: <http://aclweb.org/anthology/>, <http://aclanthology.info/>
2. *Behavior Research Methods*. URL: <https://link.springer.com/journal/13428>
3. *Computational Linguistics*. URL: <http://www.mitpressjournals.org/loi/coli>
4. *Computer Speech and Language*. URL: <https://www.journals.elsevier.com/computer-speech-and-language>
5. *International Journal of Corpus Linguistics*. URL: <https://benjamins.com/#catalog/journals/ijcl/main>
6. *Journal of Information Retrieval*. URL: <http://www.springer.com/computer/database+management+%26+information+retrieval/journal/10791>
7. *Journal of Machine Learning*. URL: <http://www.springer.com/computer/ai/journal/10994>
8. *Language and Linguistics Compass*. URL: [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1749-818X](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1749-818X)
9. *Language Resources and Evaluation*. URL: <https://link.springer.com/journal/10579>
10. *Machine Translation*. URL: <https://link.springer.com/journal/10590>
11. *Natural Language Semantics*. URL: <https://link.springer.com/journal/11050>
12. *Transactions of the Association for Computational Linguistics*. URL: <https://www.transacl.org/ojs/index.php/tacl/issue/view/13>
13. *Диалог*. URL: <http://www.dialog-21.ru/digest/>

Иные журналы, входящие в перечень ВАК/БД SCOPUS/WoS

Дополнительная учебная литература:

1. Bender E. M. Linguistic fundamentals for natural language processing: 100 essentials from morphology and syntax //Synthesis Lectures on Human Language Technologies. – 2013. – Т. 6. – №. 3. – С. 1-184.
2. Cohen K. B., Demner-Fushman D. Biomedical natural language processing. – John Benjamins Publishing Company, 2014.
3. Farzindar A., Inkpen D. Natural language processing for social media //Synthesis Lectures on Human Language Technologies. – 2015. – Т. 8. – №. 2. – С. 1-166.
4. Fitzpatrick E., Bachenko J., Fornaciari T. Automatic detection of verbal deception //Synthesis Lectures on Human Language Technologies. – 2015. – Т. 8. – №. 3. – С. 1-119.
5. Liu B. Sentiment analysis: Mining opinions, sentiments, and emotions. – Cambridge University Press, 2015.
6. Schäfer R., Bildhauer F. Web corpus construction //Synthesis Lectures on Human Language Technologies. – 2013. – Т. 6. – №. 4. – С. 1-145.
7. 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.
8. Анисимов А.В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык – Киев: Наук, думка, 1991. – 208 с.
9. Марчук, Ю.Н. Компьютерная лингвистика: учебное пособие / Ю.Н. Марчук. – М.:

АСТ: Восток-Запад, 2007. – 317 с.

10. Николаев И.С., Митренина О.В., Ландо Т.М. Прикладная и компьютерная лингвистика. – URSS, 2017. – 320 с.
11. Новое в зарубежной лингвистике: Вып. XXIV. Компьютерная лингвистика: Пер. с англ./Сост., ред. и вступ. ст. Б. Ю. Городецкого. – М.: Прогресс, 1989. – 432 с.

Ресурсы информационно-телекоммуникационной сети «Интернет»

1. Elibrary.ru: научная электронная библиотека [Электронный ресурс]. URL: http://elibrary.ru/project_risc.asp (01.09.2016).
2. Scopus: database [Электронный ресурс]. URL: <https://www.scopus.com/> (01.09.2016).
3. Web of Science: database [Электронный ресурс]. URL: <http://login.webofknowledge.com/> (01.09.2016).
4. ЭБС Ebrary [Электронный ресурс]. – URL: <http://site.ebrary.com/lib/tomskuniv/>

Перечень используемых информационных технологий, используемых при осуществлении образовательного процесса, включая программное обеспечение, информационные справочные системы:

- Описание материально-технической базы
 1. Аудитория, оборудованная проектором и компьютером с доступом в сеть Интернет.

12. Язык преподавания: русский.

13. Автор: **Шиляев Константин Сергеевич**, кандидат филологических наук, доц. кафедры общего, славяно-русского языкознания и классической филологии ТГУ.

Рецензент: **Тубалова Инна Витальевна**, доктор филологических наук, доцент кафедры общего, славяно-русского языкознания и классической филологии ТГУ.

Программа одобрена на заседании учебно-методической комиссии филологического факультета _____ года, протокол № _____.

Председатель комиссии, доцент _____ Ю.А. Тихомирова