

Паспорт расчетно-графического задания (работы)

по дисциплине «Компьютерная лингвистика», 4 семестр

1. Методика оценки

Выполнение расчетно-графического задания (работы) (далее - РГЗ(Р)) является формой текущей аттестации (контроля) по дисциплине, предусмотренной учебным планом.

Цель РГЗ(Р): студенты должны освоить и научиться программной реализации методов автоматической обработки текстовой информации в среде R.

Обязательным элементом РГЗ(Р) являются решение задач анализа текстовой информации, интерпретация и анализ полученных результатов.

Номер задания (варианта) назначается преподавателем.

РГЗ(Р) выполняется индивидуально.

Количество заданий достаточно для обеспечения, каждого обучающегося индивидуальным заданием РГЗ(Р).

Замена задания РГЗ(Р) осуществляется по согласованию с преподавателем из числа резервных (не занятых) заданий.

Перед выполнением задания студент должен ознакомиться с учебными материалами к лекционным занятиям и рекомендованной литературой.

Преподаватель осуществляет руководство по выполнению задания, оказывает консультационную помощь и принимает отчет по РГЗ(Р).

По результатам выполнения РГЗ(Р) готовится отчет, который состоит из следующих частей:

1. Титульный лист.
2. Теоретическая часть, представляющая собой описание заданных методов и алгоритмов анализа текстовой информации, особенностей и ограничений их использования.
3. Практическая (исследовательская) часть (подробное описание хода исследования в R, фрагменты кода на R с комментариями и краткими пояснениями).
4. Выводы по работе в произвольной форме (оценка рассматриваемого метода анализа текстов, удобство реализации, анализ результатов и т.п.).
5. Список использованной литературы, интернет-источников и программных средств.
6. Приложение (полный код на языке R)

Требования к оформлению:

Объем РГЗ(З) до 25 страниц машинописного текста формата А4. Шрифт Times New Roman, 12. Формулы набираются в редакторе Math Type. Размещение сканированных формул не допускается. Нумерация страниц сквозная, в нижней части листа по центру арабскими цифрами. Работа должна быть отредактирована, не содержать орфографических, синтаксических и стилистических ошибок.

Отчет в установленные сроки сдается на кафедру для проверки. Преподаватель оценивает качество работы, отмечает положительные стороны и недостатки работы и определяет, допускается ли она к защите. При необходимости преподаватель возвращает РГЗ(Р) студенту для доработки и устанавливает сроки повторного предоставления для проверки. До защиты работы студентом должны быть сделаны необходимые исправления и дополнения по всем замечаниям преподавателя.

При положительном результате оценивания РГЗ(Р) студент её распечатывает, передает на кафедру и защищает до сессии в назначенное преподавателем время.

Защита РГЗ(Р) состоит в индивидуальном устном собеседовании студента с преподавателем. В процессе защиты выявляется уровень знаний студента, степень его самостоятельности при выполнении работы. По результатам защиты студенту выставляется оценка в соответствии с критериями, приведенными в п. 2 настоящего Паспорта.

2. Уровни сформированности компетенций и критерии оценки

РГЗ(Р) считается выполненной **на продвинутом** уровне, если все необходимые расчеты произведены самостоятельно и без ошибок; все разделы РГЗ(Р) выполнены правильно и в полном объеме; работа оформлена в соответствии с требованиями; сдана преподавателю в указанные сроки и не возвращалась для доработки; даны полные и развернутые выводы и рекомендации; на защите студентом даны уверенные и аргументированные ответы, что свидетельствует об уровне сформированности компетенций и соотнесенных с ними индикаторов на продвинутом уровне. Оценка составляет *от 36 до 40 баллов*.

РГЗ(Р) считается выполненной **на базовом** уровне, если все необходимые расчеты произведены самостоятельно и без существенных ошибок; все разделы РГЗ(Р) выполнены правильно, но есть замечания к полноте предоставления информации; работа оформлена в соответствии с требованиями; сдана преподавателю в указанные сроки и однократно возвращалась студенту для незначительной доработки; в заключении даны выводы и рекомендации; на защите студентом допущены не принципиальные ошибки, что свидетельствует об уровне сформированности компетенций и соотнесенных с ними индикаторов на базовом уровне. Оценка составляет *от 28 до 35 баллов*.

РГЗ(Р) считается выполненной **на пороговом** уровне, если все необходимые расчеты произведены самостоятельно, но с ошибками, часть из которых носит принципиальный характер; есть замечания к полноте предоставления информации; работа оформлена в соответствии с требованиями; сдана преподавателю, но неоднократно возвращалась студенту для доработки; в заключении даны краткие выводы; защита РГЗ(Р) вызывает у студента серьезные затруднения, что свидетельствует об уровне сформированности компетенций и соотнесенных с ними индикаторов на пороговом уровне. Оценка составляет *от 20 до 27 баллов*.

РГЗ(Р) считается **не выполненной**, если расчеты произведены с серьезными ошибками; есть замечания к полноте предоставления информации и оформлению; РГЗ(Р) была сдана преподавателю, но неоднократно возвращалась студенту для доработки, что не привело к улучшению ее качества; РГЗ(Р) не допущена до защиты, что свидетельствует об не сформированности компетенций и соотнесенных с ними индикаторов. Оценка составляет *менее 20 баллов*.

3. Шкала оценки

В общей оценке по дисциплине баллы за РГЗ(Р) учитываются в соответствии с правилами балльно-рейтинговой системы, приведенными в рабочей программе дисциплины.

РГЗ(Р) как форма текущей аттестации (контроля) по дисциплине считается

успешно выполненной, если сумма полученных баллов по всем его заданиям составляет от 20 до 40 баллов включительно.

4. Примерный перечень тем (вариантов) РГЗ(Р)

Темы РГР по дисциплине: «Компьютерная лингвистика»

1. Анализ взаимосвязи слов в тексте на основе n-грамм.
2. Форматы хранения текстовой информации в R, преобразование форматов, анализ частоты слов в документе на основе статистики tf-idf.
3. Классификация текстов с помощью Байесовского алгоритма и алгоритма Роккио, оценка качества классификации, сравнение методов.
4. Классификация текстов с помощью метода наименьших квадратов и метода опорных векторов, оценка качества классификации, сравнение методов.
5. Классификация текстов с помощью алгоритмов деревьев решений (выбрать два разных алгоритма), оценка качества классификации, сравнение методов.
6. Классификация текстов с помощью алгоритма k-ближайших соседей и алгоритма Роккио, оценка качества классификации, сравнение методов.
7. Кластеризация текстов с помощью иерархических методов (выбрать два разных метода), оценка качества кластеризации, сравнение методов.
8. Кластеризация текстов с помощью алгоритма k-средних и нечеткого алгоритма c-средних, оценка качества кластеризации, сравнение методов.
9. Кластеризация текстов с помощью плотностного алгоритма DBSCAN, оценка качества кластеризации.
10. Кластеризация текстов с помощью инкрементного алгоритма S^2ICM , оценка качества кластеризации.
11. Кластеризация текстов с помощью нейросетевого алгоритма SOM, оценка качества кластеризации.
12. Синтаксические анализаторы, принципы построения, примеры анализаторов
13. Морфологические и синтаксические анализаторы в поисковых системах, принципы работы, примеры систем
14. Сетевые модели представления знаний для компьютерной обработки, примеры моделей.
15. Фреймовые модели представления знаний для компьютерной обработки, примеры моделей.
16. Онтологические модели представления знаний для компьютерной обработки, примеры моделей.

Допускается выбор собственной темы по согласованию с преподавателем.

Задание на РГР:

1. Описать заданный метод анализа текстовой информации (задача, для решения которой разработан метод, предварительная обработка текста для использования метода, основные формулы, алгоритм применения метода, достоинства, недостатки, вычислительная сложность метода и т.д.).
2. Выбрать текстовые документы (корпуса текстов), которые будут использоваться в качестве исходных данных.
3. Реализовать решение задачи анализа текста в среде R.

4. Сделать выводы по работе в произвольной форме (оценка рассматриваемого метода анализа текстов, удобство реализации, анализ результатов и т.п.).
5. Подготовить отчет по РГЗ(Р).