О первой практической работе

Главной практической задачей первого блока курса «Автоматический анализ текстов» является самостоятельная постановка эксперимента по теме «Классификация текстов на основе стилометрических характеристик». Для вашего удобства эксперимент предложено разделить на этапы и выполнять их последовательно в течение двух с половиной месяцев.

В самом начале вам предстоит определиться с направлением исследований. Вы можете выполнять классификацию текстов по жанру, авторству, стилю и тд (всегда можно придумать своё направление классификации, но перед этим посоветуйтесь с преподавателем).

Результатом выполнения задания будет комплексная программа, выполняющая предобработку, классификацию и верификацию, а также отчёт, в котором вы изложите все особенности эксперимента, и небольшая презентация, с помощью которой вы покажете результаты работы. Код, отчёт и презентацию необходимо загрузить на мудл отдельным архивом в день защиты.

1 этап (после лекции о предобработке)

Определитесь с направлением будущих экспериментов и подготовьте корпус. Корпус должен содержать не менее 90 текстов. Объём одного текста – не менее 30 предложений. Помните: ваш корпус должен быть сбалансированным и должен содержать примерно одинаковое количество элементов каждого класса.

Следующий шаг – предобработка. В соответствии с задачей и темой исследования сформулируйте и реализуйте требования к оформлению и разметке текстов. Отдельно подумайте над тем, что и почему вы будете считать шумом.

По желанию вы можете выполнять этот этап в команде (не более 3 человек).

В отчёте отразите выбранную тему, а также правила, по которым формировался и оформлялся корпус.

2 этап (после лекции о стилометрии)

Выделите стилометрические характеристики, необходимые для решения задачи. Соберите алгоритмы для их вычислений и запишите результат в качестве вектора.

Перечислите в отчёте используемые характеристики и обоснуйте их выбор.

3 этап (после лекции о верификации)

Выберите не менее 3 классификаторов и проведите эксперименты по классификации текстов по выбранному вами ранее направлению.

Оцените результат классификации и выберите лучший классификатор.

В отчёте расскажите о выбранных классификаторах и их работе. Проанализируйте полученные результаты и сделайте выводы о проведённых экспериментах.