

ПОДХОД К ОПРЕДЕЛЕНИЮ СЛОЖНОСТИ ДИСЦИПЛИН

Аннотация. В исследовании проанализированы существующие способы для определения сложности текстов. В статье описаны подходы к вычислению сложности как для реализованных и изучаемых элективных дисциплин, так и для новых. Результатом работы является разработанный алгоритм, позволяющий вычислить сложность элективов с учетом параметров, определенных на основе данных Тюменского государственного университета.

Ключевые слова: сложность текста, электив, параметр сложности, кластеризация, метод Саати.

Введение. В настоящее время существуют различные варианты определения сложности текста, наиболее распространенным является комбинирование количественных и качественных параметров. В качестве факторов, влияющих на сложность, обычно выступают удобочитаемость, лексическое разнообразие и длина слов.

В статьях [1-3] представлены результаты проведенных исследований влияния различных параметров на сложность слов и учебных текстов. Исходя из приведенных работ, были сделаны выводы: выявлена прямая зависимость индекса Флеша-Кинкейда со сложностью, а высокий показатель читабельности, частотность, степень абстрактности и многие качественные параметры совместно не всегда формируют более высокую сложность учебных текстов.

Майеру Р.В. принадлежит множество работ, посвященных исследованию и совершенствованию методов оценки дидактической сложности учебных текстов [4-5]. Он, как и многие другие авторы, считает, что сложность текста вычисляется с учетом нескольких параметров, а именно текстовой составляющей, общей сложности рисунков и формул, включенных в текст. Общая сложность зависит от объема, информативности текста, средней сложности предложений, числа изображенных объектов и связей на рисунках, а также количеством формул и средним числом символов в них.

Проблема исследования. Ныне существующие образовательные модели направлены на индивидуализацию процесса обучения. В течение всего периода обучения в учебном заведении обучающийся строит собственную траекторию, состоящую из основных и дополнительных дисциплин. Выбор элективов не всегда прост: так студенты часто сталкиваются с различными проблемами: трудность выбора предметов с учетом различных критериев, невозможность записаться на желаемую дисциплину и другими.

В настоящее время имеются реализации подбора элективов, учитывающих интересы, предпочтений, а также профстандартов, однако остается открытым вопрос: как подобрать элективы с учетом предпочтений и принимать во внимание сложность подбираемых дисциплин, поэтому целью данного исследования является разработка алгоритма вычисления сложности элективов для различных направлений подготовки.

Материалы и методы. Для проведения исследования в качестве входных данных были взяты вопросы для аттестации из файлов оценочных материалов по дисциплинам. Для определения основных терминов из вопросов к промежуточной аттестации были выполнены:

- 1) предобработка текста, которая состоит из приведения к нижнему регистру, нормальной форме слов и удаления стоп-слов;
- 2) применение метода Yake для извлечения ключевых слов.

Проанализировав имеющиеся факторы, влияющие на освоение дисциплин, было принято решение о формировании алгоритма определения сложности элективных дисциплин (подход 1), зависящего от направления подготовки студента, семестра изучения дисциплины, оценок по изученным дисциплинам, отзывов, а также лексической сложности.

Первым шагом определения сложности элективов, является кластеризация дисциплин, что позволит получить набор возможных областей знаний, состоящих из близких по тематике учебных предметов. Выбор метода кластеризации был основан на результатах исследований показателей эффективности научных публикаций, которые показали превосходство KMeans во многих работах, поэтому было принято решение об использовании метода KMeans. Для определения оптимального количества кластеров использован метод локтя и метод силуэтов.

Для определения весовых значений коэффициентов в данной работе применяется метод парных сравнений, а именно метод Саати, суть которого состоит в указании степени важности одной альтернативы относительно другой, выполняемые одним экспертом. На основе статьи [6] использовался модифицированный метод оценки от 1 до 5, он более удобен и характеризуется большей степенью уверенности эксперта. В итоге для каждого критерия был получен соответствующий ему вес (табл. 1).

Таблица 1

Весовые коэффициенты параметров сложности элективов

	<i>Весовой коэффициент</i>
Направление подготовки	0,25
Семестр	0,19
Результаты ПА	0,12
Оценка легкости по Отзывус	0,06
Лексическая сложность (частотность терминов)	0,38

На основе имеющихся данных и полученных весовых значений для каждого электива направлений подготовки Школы компьютерных наук была вычислена сложность.

Однако, ежегодно в базу добавляются новые элективы, отзывов о которых не может существовать, также как и результатов аттестации, поэтому значение сложности таких дисциплин можно определять только по текстам оценочных материалов (подход 2). Для определения параметров, на основе которых будет определяться сложность для них, было решено вычислить корреляцию между сложностью и наиболее популярными показателями сложности текстов.

Исходя из значений, полученных при создании матрицы корреляции между параметрами сложности (уровень легкости по Отзывус, ttr, msttr, средняя частота терминов, доля абстрактных слов в тексте) и вычисленной лексической сложностью (см. рис. 1), было принято решение вычислять сложность как среднее между ttr (type-token ratio), частотностью терминов и долей абстрактных слов.

Результаты исследования. Во время исследования были рассмотрены подходы к определению сложности элективных дисциплин, разработан и апробирован алгоритм определения сложности элективной дисциплины. В ходе кластеризации дисциплин было получено 11 основных областей знаний — кластеров, каждый из которых был разбит на более узкие предметные области.

На основе изученных работ и данных Тюменского государственного университета был определен набор параметров, от которых зависит сложность элективных дисциплин. С помощью метода Саати получены весовые коэффициенты формулы вычисления сложности, для 10 образовательных направлений подготовки Школы компьютерных наук была определена сложность более 400 элективов.

	ease_rate_otz	ttr	msttr	freq	abstract	text_complexity
ease_rate_otz	1.000000	0.025000	-0.027000	0.031000	0.043000	-0.822000
ttr	0.025000	1.000000	0.151000	0.333000	0.017000	0.236000
msttr	-0.027000	0.151000	1.000000	-0.155000	-0.121000	0.061000
freq	0.031000	0.333000	-0.155000	1.000000	-0.010000	0.048000
abstract	0.043000	0.017000	-0.121000	-0.010000	1.000000	0.019000
text_complexity	-0.822000	0.236000	0.061000	0.048000	0.019000	1.000000

Рис. 1. Матрица корреляции параметров сложности и вычисленной сложностью

Заключение. В результате проведенной работы был разработан алгоритм, с помощью которого для различных направлений подготовки можно вычислить сложность элективных дисциплин с учетом выделенных параметров. Данное решение предусматривает вариант определения сложности для новых дисциплин, при отсутствии информации о результатах промежуточной аттестации и отзывов, основываясь исключительно на параметрах лексической сложности текстов.

СПИСОК ЛИТЕРАТУРЫ

1. Мартынова Е.В., Солнышкина М.И., Мерзлякова А.Ф., Гизатулина Д.Ю. Лексические параметры учебного текста (на материале текстов учебного корпуса русского языка) / Е.В. Мартынова, М.И. Солнышкина, А.Ф. Мерзлякова, Д.Ю. Гизатулина — Текст: непосредственный // Филология и культура. — 2020 — № 3 (61). — С. 72-80.
2. Морозов Д.А., Иомдин Б.Л. Критерии семантической сложности / Д.А. Морозов, Б.Л. Иомдин — Текст: электронный — URL: <https://publications.hse.ru/mirror/pubs/share/direct/398187895.pdf> (дата обращения: 20.02.2024).
3. Вахрушева А.Я. Оценка лексической и синтаксической сложности текстов, представленных в российских учебных пособиях / А.Я. Вахрушева — Текст: непосредственный // Скориновские чтения. — 2023. — С. 373-387.
4. Майер Р.В. Дидактическая сложность учебных текстов и ее оценка / Р.В. Майер — Текст: электронный — URL: http://maier-rv.glazov.net/Mayer_monograph2020.pdf (дата обращения: 10.02.2024)
5. Майер Р.В. Об оценке сложности и информативности учебных текстов / Р.В. Майер — Текст: непосредственный // Территория науки. — 2018. — № 6. — С. 28-33.
6. Тихомирова А.Н., Сидоренко Е.В. Модификация метода анализа иерархий Т. Саати для расчета весов критериев при оценке инновационных проектов / А.Н. Тихомирова, Е.В. Сидоренко. — Текст: непосредственный // Современные проблемы науки и образования. — 2012. — № 2. — С. 261.
7. Crossley S.A., Cobb T., & McNamara D.S. Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. — Текст: электронный — URL: <https://sci-hub.ru/10.1016/j.system.2013.08.002> (дата обращения: 18.02.2024).