

Ю. Ф. Шпаковский, БГТУ¹, 2003

ФОРМУЛЫ ЧИТАТЕЛЬНОСТИ² КАК МЕТОД ОЦЕНКИ КАЧЕСТВА КНИГИ

Рассмотрены основные этапы разработки формулы читабельности. Анализируются методы определения трудности текста.

The main sages of construction of readability formula are considered. Methods of determination of text difficulty are analyzed.

1. Постановка проблемы. Прежде чем оценивать качество книги, необходимо понять, в чем ее сущность. Исследователи считают, что «сущность книги — в фиксации смысловой информации с целью ее передачи во времени и пространстве» [1]. Далее возникает вопрос: что такое информация? До сих пор всеми принятого определения данного понятия не существует. Это одно из первичных, неопределяемых понятий науки [2]. Поэтому под словом «информация» будем понимать концептуально связанные между собой сведения, данные, изменяющие наши представления о явлении или объекте окружающего мира. Ю. А. Шрейдер считает, что наше интуитивное понятие об информации можно выразить в терминах взаимодействия объектов произвольной физической природы [3]. Имеется два объекта A и B , для каждого из которых заданы множества возможных состояний и определены состояния в каждый момент времени. Для пары взаимодействующих объектов можно ввести промежуточный объект S , через который передается взаимодействие, например: два электрона и электромагнитное поле, автор и читатель книги и сама книга и т. д. Если объект S заменить системой с существенно иными энергетическими и массовыми характеристиками, то в ряде случаев динамика поведения объектов A и B не изменится. Это подтверждает тот факт, что для описания взаимодействия наиболее существенны информационные характеристики.

¹ Белорусский государственный технологический университет, кафедра редакционно-издательских технологий.

² От английского слова «readability».

Из всего вышесказанного можно сделать следующие выводы:

1. Книга — это единство содержания (информации) и материальной формы, однако первостепенное значение для оценки ее качества должна иметь ее семантическая сторона. Данное положение не умаляет значения полиграфического исполнения книги и не снижает требований, предъявляемых к ее оформлению. Но не следует забывать, что цель любого издания — это, прежде всего, передача информации.

2. Книга есть промежуточное звено между отправителем информации и ее потребителем. И качество книги зависит от того, насколько изменилось поведение потребителя информации (другими словами, насколько изменилось представление читателя об окружающем мире).

Таким образом, оценка качества книги может сводиться к тому, чтобы узнать, насколько изменился тезаурус³ читателя после прочтения книги. Так мы узнаем, какое количество семантической информации получил читатель. Однако, получить такую характеристику очень сложно, так как можно лишь весьма ориентировочно оценить запас сведений, которым обладает человек до и после прочтения какого-либо текста. Кроме набора тестов, с помощью которого тестируют каждого члена группы, предложить что-то еще не представляется возможным. Данный метод не используется ввиду большого объема процедуры.

Поэтому в данной статье предлагаются пути косвенной оценки количества информации, которую может получить читатель при чтении какого-либо текста. В этом случае используются формулы читабельности, с помощью которых в количественном виде можно определить трудность того или иного текста.

2. Анализ последних исследований. Исследования по читабельности начались в 20-х гг. прошлого столетия в США и до сих пор интерес к данной проблеме не ослабевает. В 1928 г. М. Вогель и К. Вашбурн предложили общую схему исследования читабельности [4], которая была взята за основу во всех последующих работах. Она сводилась к нахождению объективных характеристик текста, которые достаточно высоко коррелируют с численной оценкой результата тестирования, и составлению уравнения регрессии между оценкой

³ Некоторый список объектов, свойств и классов объектов с определенными связями между ними [3].

тестирования, с одной стороны, и оценками параметров текста — с другой. Дальнейшие исследования были направлены на улучшение и упрощение формул читабельности. В настоящее время в уравнении регрессии используется 2—3 фактора трудности. Наибольшей популярностью в США пользуются формулы читабельности Р. Флеша [5], Э. Фрая [6, 7], Дале-Чел [8] и др. Обработка материала сегодня производится на компьютере с помощью таких программ, как: Readability Calculations, Intext (анализирует текст на английском, немецком, французском, испанском, голландском, датском и шведском языках), Nisus Writer, MS Word, Corel WordPerfect и др.

В СССР проблемами читабельности занимались, но не так активно, как в США. Известны диссертации М. С. Мацковского [9] и Я. А. Микка [10, 11], которые разработали формулы читабельности для русского и эстонского языков соответственно на основе научно-популярных текстов. Отдельные статьи опубликовали Е. Б. Колмен [12], М. Крسمанович [13], Ю. Тулдава [14], А. Д. Матеева [15].

3. Цель работы — описать основные этапы разработки формулы читабельности как метода оценки качества изданий с анализом прделанной работы по данной теме.

4. Изложение основного материала. Следует отметить, что под словом «читабельность» Дэйла и Чалл предложили понимать некоторую характеристику печатного материала, зависящую от всех элементов внутри данного материала, которые влияют на успешность его «усвоения» определенной группой читателей. Мерой такого успешного усвоения является то, насколько средний читатель интересующей нас группы понимает исследуемый материал, в какой мере скорость, с которой он его читает, приближается к оптимальной, и, наконец, какой интерес представляет данный материал для этого среднего читателя [16]. В современных словарях дается более упрощенное определение: «Читабельность — легкость понимания вследствие стиля изложения» [17].

Так как методы прямой оценки количества информации практически не применяются из-за своей трудоемкости, предлагаем использовать косвенную оценку того количества информации, которую может получить читатель при чтении текста.

Прежде согласимся с автором работы [18], который выдвигает следующие положения:

1. Чем благоприятнее условия для максимально возможного изменения тезауруса, тем больше будет степень этого изменения и, следовательно, тем больше семантической информации содержит данный текст относительно данного тезауруса.

2. Необходимыми, хотя и недостаточными условиями того, чтобы произошло наибольшее изменение тезауруса при действии на него текста, являются:

а) текст должен быть адекватно понят читателем;

б) уровень подготовленности читателя не должен на много превосходить тот уровень, на который рассчитан текст.

Из этого следует, что наибольшее количество информации читатель получает в том случае, когда уровень его знаний соответствует тому уровню, на который рассчитан этот текст (при этом разумно учитывать лишь адекватно воспринятую информацию). Таким образом, повысить качество издания (для того, чтобы читатель получил максимальное количество информации) можно лишь, добившись оптимального сочетания сложности текста и уровня подготовленности читателя. Разработка формулы читабельности позволит определить, насколько тот или иной текст соответствует тому или иному читателю. Кроме этого будут выявлены основные *компоненты сложности текста*⁴. Таким образом, разработанная формула читабельности будет являться одной из вполне определенных оценок качества книги для конкретного читателя.

Итак, основные этапы разработки формулы читабельности следующие:

1) нахождение методов для определения трудности различных текстов для данной группы лиц;

2) выбор структурных элементов (и только тех, которые поддаются точному измерению) всех исследуемых текстов;

3) нахождение некоторой функции, которая бы однозначно описывала зависимость между результатами на понимание печатного материала интересующей группы лиц и выбранными структурными элементами текста.

Этап 1. Материалом для эксперимента послужили отрывки из учебных изданий для вузов по разделам химии [20, 21, 22, 23, 24]. Длина выбранных отрезков варьировалась от 100 до 300 слов: доста-

⁴ Характеристики (свойства) текста, от которых зависит трудность его понимания и усвоения [19].

точно длинные, чтобы выявить все структурные элементы текста, которые влияют на его понимание, и достаточно короткие, чтобы не затруднять анализ. Отрывки могли быть знакомы испытуемым (студенты 3, 4, 5-го курсов БГТУ), так как данные учебники используются в процессе обучения в вузе. Однако, по нашему мнению, это вряд ли исказит результаты по двум причинам:

а) выборка была репрезентативной;

б) если испытуемому знакомы все тексты, он не помещает их в одну категорию (например, легкий), так как языковые параметры текстов все же различны, что влияет на усвоение материала. И даже в том случае, если испытуемому предложить трудный текст, но уже знакомый, и легкий незнакомый, он субъективно оценивает первый текст как более трудный. Это, по-видимому, связано с тем, что при чтении человек совершает определенные мыслительные операции. При чтении уже знакомого текста все равно приходится каждый раз осмысливать слова, фразы, предложения. И более длинные слова и предложения заставляют совершать больше мыслительных операций, и текст кажется более трудным. В целом тесты проводились по пройденному материалу, в подготовке к которому использовались различные учебные издания.

Для того, чтобы исследовать влияние внетекстовых элементов (таблиц, формул, иллюстраций) на понимание материала были выбраны отрывки с соответствующими структурными элементами и переработаны эти же отрывки с заменой внетекстовых элементов описанием в словесном виде.

Всего было отобрано 28 отрывков по следующим признакам: средняя длина предложения (в словах), средняя длина слова в буквах, процент фраз в повторениях, количество терминов, количество абстрактных/конкретных существительных. Исключались те тексты, значения признаков которых совпадали или были близки к средним.

Будем различать *трудность текста*, которая устанавливается по результатам понимания данного текста (т. е. экспериментально), и *сложность текста*, которая определяется при помощи анализа этого текста, например по проценту незнакомых слов, по длине предложений, по сложности логической структуры и другим компонентам сложности [19].

В исследованиях читабельности трудность текста определяется различными методами: постановка вопросов к тексту, сводка основного содержания текста, методика дополнения, экспертные оценки испы-

туемых о трудности текста, составление плана или схем текста, угадывание текста по буквам, интонирование, пересказ, скорость чтения текста [19, 25, 26].

В нашем исследовании использовалось два метода: методика дополнения и экспертные оценки трудности текста.

Методика дополнения — это заполнение пропусков в тексте, в котором слова через определенный интервал (в английском языке, например, каждое пятое слово) заменены точками. Плюсы данной методики состоят в том, что пропускается всегда только одно слово, и слова пропускаются не по собственному усмотрению исследователя, а по строгому правилу.

Используя данную методику, необходимо было ответить на три вопроса: 1) Сколько слов оставлять между пропусками? 2) Оценивать ли синонимы к пропущенным словам правильными ответами или нет? 3) Какой длины должна быть методика дополнения?

Отвечая на первый вопрос, мы должны иметь в виду следующее: заполнение одного пропуска не должно зависеть от заполнения другого, только в этом случае каждый пропуск будет представлять самостоятельное задание. Это условие и определяет количество слов между пропусками, которое будет различным как для отдельных языков, так, по всей видимости, и для разных видов текста (научный, научно-популярный, художественный и т. д.). Для английского языка пропускается каждое пятое слово, для эстонского — седьмое.

Для того чтобы выяснить, сколько слов необходимо оставлять между пропусками в нашем случае, нами был поставлен эксперимент. Испытуемым предлагали пропуск с тремя словами по обе стороны пропуска. Если он не мог правильно заполнить пропуск, то увеличивали количество слов до четырех с обеих сторон. И так до того момента, пока испытуемый не давал правильного ответа. В эксперименте приняло участие 25 студентов. В итоге мы получили результат — пропускать необходимо каждое 8-е слово.

Второй вопрос мы решили таким образом. Известно, что ученые, работающие на базе английского языка, правильными считают только те слова, которые употребил автор. В отечественной практике используется также вариант, когда правильными считаются все ответы, подходящие по смыслу [27, 28]. В этом случае корреляция составляет 0,85 (в первом — 0,67). В дальнейшем при обработке материала второй способ удобен тем, что выборка текста намного меньше, хоть и

времени на оценку результатов понадобится чуть больше. Поэтому мы решили использовать второй вариант.

Решение третьей проблемы сводится к следующему. В каждом отрывке есть легко и трудно восстанавливаемые слова. Соотношение этих слов должно быть примерно одинаковым в методике дополнения и во всем тексте. Оптимальное количество пропускаемых слов можно установить по следующей формуле [29]:

$$n_0 = \left(\frac{t\sigma}{\Delta X} \right)^2,$$

где t — константа, которая определяется желаемым уровнем вероятности (при 95 %-ном уровне вероятности $t = 1,96$);

σ — среднее квадратическое отклонение;

ΔX — предельная ошибка выборки.

По этой же формуле можно определить и количество испытуемых, которые должны участвовать в проведении эксперимента.

После решения всех перечисленных вопросов к каждому отрывку была составлена инструкция следующего вида:

1. Укажите номер теста.
2. Зафиксируйте время начала работы с текстом.
3. В данном отрывке текста пропущено каждое 8-е слово. Запишите на листы ответов слова, которые больше всего подходят в эти пропуски.
4. Зафиксируйте время окончания работы с текстом.

Пропуски в текстах были пронумерованы, и испытуемые записывали на листах ответов номер пропуска, который они заполнили. Кроме этого, было зафиксировано время работы с текстом, что дополнительно позволит использовать и этот метод определения трудности текста.

Экспертные оценки трудности текста. Данная методика также дает хороший коэффициент корреляции [28, 30]. В данном случае индекс трудности текста определяется на основе экспертных оценок студентов. Суть метода заключалась в следующем: после прочтения отрывка испытуемым предлагалось оценить его трудность по шестибальной шкале:

- 1 — очень легкий текст;
- 2 — легкий текст;
- 3 — текст со средней трудностью;

- 4 — трудный текст;
- 5 — очень трудный текст;
- 6 — сверхтрудный текст.

Для того, чтобы исключить поверхностное знакомство испытуемых с текстом, что исказило бы результаты при оценке его трудности, студентам перед суждением о трудности текста по шкале предлагалось выписать несколько ключевых слов и выразить основное содержание отрывка одним простым предложением. Эти меры заставили испытуемых тщательно изучать тексты. Кроме этого, мы просили подчеркнуть в отрывках все незнакомые слова.

По данным этих экспериментов будут найдены показатели трудности текста.

Этап 2. На данном этапе необходимо вычислить значения признаков текста. Выше было сказано, что должны быть выбраны такие структурные элементы, которые поддаются точному измерению. Это выполнимо только в случае, когда известны единицы текста. Они должны отвечать двум основным требованиям [31]:

- а) быть простыми в установлении;
- б) не должны составлять самого себя.

Этим требованиям отвечает общепризнанная единица — слово⁵. Выбор остальных единиц вызывает определенную трудность. Например, выделение таких структурных элементов как слово с определениями (выражает понятие) или термин суждения требует детального семантического анализа текста. Поэтому вслед за Я. А. Микком [31] выделим еще одну единицу — фразу. Фраза есть сказуемое со всеми словами, принадлежащими к нему, в том числе и с группой подлежащего⁶. Таким образом, фразой является простое предложение, главная и придаточная части сложноподчиненного предложения (две фразы), каждая часть сложносочиненного предложения.

На основе вышесказанного были выделены следующие признаки текста:

- 1—3. Длина абзаца во фразах/словах/печатных знаках.
- 4—6. Длина предложения в словах/фразах/печатных знаках.

⁵ В тексте слова четко разграничены (набор букв между пробелами) и являются целостным средством выражения простого понятия.

⁶ Форма глагола действительно является хорошим средством установления единицы текста.

7—9. Длина самостоятельного предложения⁷ во фразах/словах/печатных знаках.

10—12. Длина фразы в словах/буквах/печатных знаках.

13—14. Длина слова в буквах/печатных знаках.

15. Длина текста в печатных знаках.

16. Количество упоминаний личностей на 100 слов.

17. Конкретных имен существительных на 100 имен существительных.

18. Абстрактных имен существительных на 100 имен существительных.

19. Средняя абстрактность имен существительных.

20. Процент разных незнакомых слов.

21. Процент незнакомых слов и др.

Этап 3. На данном этапе с помощью факторного анализа из всех показателей трудности выбираются наилучшие и вычисляются значения суммарного показателя трудности текстов. После этого, вычисляются и анализируются с помощью корреляционного анализа связи между признаками и трудностью текстов. В конце составляется формула читабельности.

Выводы. Исследования по разработке формул читабельности имеют большое практическое и теоретическое значение. Разработанные формулы дают возможность оценить качество книги для определенного читателя, что позволит редакционно-издательским работникам повысить качество изданий.

⁷ Простые предложения, части сложносочиненного предложения и сложноподчиненное предложение в целом.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Баренбаум И. Е. Функциональный подход и его применение в книговедении. — В кн.: Книга и социальный прогресс. М.: Наука, 1986, с. 122—131.
2. Информатика / Под ред. К. В. Тараканова. М.: Книга, 1986.
3. Шрейдер Ю. А. О количественных характеристиках семантической информации // НТИ. — 1963. — № 10. — С. 33—48.
4. Vogel M., Washburne C. An objective method of determining grade placement of children's reading material // Elementary School Journal. — 1928. — Vol. 28. — Pp. 373—382.
5. Flesch R. The art of readable writing. N. Y., 1968.
6. Fry E. B. Fry's readability graph: Clarifications, validity, and extensions to level 17 // Journal of reading. — 1977. — Vol. 21. — Pp. 242—252.
7. Fry E. B. Readability versus leveling // The reading teacher. — 2002. — Vol. 56. — № 3. — Pp. 286—291.
8. Chall J. S., Dale E. Manual for the new Dale-Chall readability formula. Cambridge, MA: Brookline, 1995.
9. Мацковский М. С. Проблема понимания читателями печатных текстов (социологический анализ). Автореф. канд. дис. — М.: НИИ СИ АН СССР, 1973. — 32 с.
10. Микк Я. А. О факторах понятности учебного текста. Автореф. канд. дис. — Тарту: ТГУ, 1970. — 24 с.
11. Микк Я. А. Теория измерения и оптимизации степени сложности учебного материала в общеобразовательной школе. Автореф. докт. дис. — М.: НИИ ОП АПН СССР, 1982. — 36 с.
12. Колмен Е. Б. Экспериментальное исследование читабельности. — «Методика преподавания иностранных языков за рубежом». М., 1976, вып. 2.
13. Крсманович М. Методы исследования удобочитаемости учебных текстов. — «Проблемы школьного учебника». М., 1974, вып. 2.
14. Тулдава Ю. Об измерении трудности текстов. — «Ученые записки Тартуского университета». 1975, вып. 345, IV, Труды по методике преподавания иностранных языков, с. 102—120.
15. Матеева А. Д. О задаче прогнозирования трудности понимания текста по объективным языковым характеристикам. — Сборник научных трудов МГПИИЯ, 1977, вып. 124, с. 60—69.
16. Chall J. S. Readability. An appraisal of research and application. Oh., 1958.
17. Weaver B. M. Leveling books K-8: Matching readers to text. Newark, DE: International Reading Association, 2000.
18. Мацковский М. С. Применение формул читабельности для получения некоторых количественных характеристик семантической информации // НТИ. Сер. 2. — 1969. — № 6. — С. 3—7.

19. Микк Я. А. Оптимизация сложности учебного текста: В помощь авторам и редакторам. М.: Просвещение, 1981. — 119 с.
20. Пилипенко А. Т., Пятницкий И. В. Аналитическая химия: В двух книгах: кн. 1 — М.: Химия, 1990. — 480 с.: ил.
21. Карапетьянц М. Х., Дракин С. И. Общая и неорганическая химия: Учеб. для вузов. — 2-е изд., перераб. и доп. — М.: Химия, 1992. — 592 с.: ил.
22. Фролов Ю. Г. Курс коллоидной химии (Поверхностные явления и дисперсные системы): Учеб. для вузов. — М.: Химия, 1982. — 400 с.: ил.
23. Стромберг А. Г., Семченко Д. П. Физическая химия: Учеб. для хим.-технол. спец. вузов / Под ред. А. Г. Стромберга. — 2-е изд., перераб. и доп. — М.: Высш. шк., 1988. — 496 с.: ил.
24. Щербина А. Э., Матусевич Л. Г., Сенько И. В., Звонок А. М. Органическая химия. Реакционная способность основных классов органических соединений: Учеб. пособие для студ. хим.-технол. спец. — Мн.: БГТУ, 2000. — 624 с.: ил.
25. Фрумкина Р. М. Понимание текста в условиях ограниченного знания словаря // НТИ. — 1965. — № 4. — С. 44—48.
26. Бормут Дж. Методика дополнения. — В кн.: Советская педагогика и школа. Тарту, 1974, вып. 9, с. 65—77.
27. Рапопорт И. А., Гохлернер М. М., Сельг Р., Соттер И. О диагностических функциях тестовой методики дополнения. — Иностранные языки в школе, 1976, № 2, с. 31—37.
28. Микк Я. А. Методика измерения трудности текста. — Вопросы психологии, 1975, № 3, с. 147—155.
29. Теория статистики: Учебник / Под ред. проф. Р. А. Шмойловой. — М.: Финансы и статистика, 1996. — 464 с.: ил.
30. Методика исследования восприятия информации. Сб. научных трудов / Под ред. Б. М. Фирсова. — Л.: НИИ ООВ АПН СССР, 1972. — 152 с.
31. Микк Я. А. Методика разработки формул читабельности. — В кн.: Советская педагогика и школа. Тарту, 1974, вып. 9, с. 78—163.