

Екатерина Пиджакова, Татьяна Казанцева'

Ekaterina Pidzhakova, Tatiana Kazantseva

ОЦЕНКА РАЗМЕРА ИНФОРМАЦИИ В СТАНДАРТАХ

ASSESSMENT OF THE SIZE OF INFORMATION IN STANDARDS

Уральский федеральный университет имени первого Президента России Б. Н. Ельцина,
г. Екатеринбург

Ural Federal University named after the first President of Russia B. N. Yeltsin, Ekaterinburg

В статье рассматривается алфавитный подход измерения размера информации в стандартах. Представлен сравнительный анализ размера информации в стандартах раздела 77.120.01 «Цветные металлы в целом» в зависимости от временного периода их принятия.

The article discusses an alphabetical approach to measuring the size of information in standards. A comparative analysis of the size of the information in the standards of section 77.120.01 «Non-ferrous metals in general» is presented, depending on the time period of their adoption.

Ключевые слова: стандарт, размер информации, единица информации, алфавитный подход, мощность алфавита

Key words: standard, amount of information, unit of information, alphabetic approach, alphabet power

Понятие «информация» достаточно широко используется в обычной жизни современного человека, термин происходит от латинского «information», что означает разъяснение, осведомление, изложение. В современном понимании термин «информация» можно трактовать как сведения, снимающие неопределенность об окружающем мире. Эти сведения являются объектом хранения, преобразования, передачи и использования [1].

По форме представления информация может быть символьная, текстовая, в которой смысл закладывается в сочетании символов, графическая и звуковая.

Национальный фонд стандартов содержит в настоящее время ~ 35000 стандартов. Это ценнейшая информация, накопленная обществом в процессе своего развития. Абсолютно все согласны, что стандарты содержат огромный объем информации, однако в современных

' Пиджакова Е. – студент магистратуры
Казанцева Т. – старший преподаватель

условиях необходимо более конкретно иметь возможность оценить размер такой информации [2].

Термин «количество информации» в соответствии с ГОСТ 8.417–2002 «ГСИ. Единицы величин» рассматривается как единица информации в двоичной системе счисления (двоичная единица информации) применительно к устройствам цифровой обработки и передачи информации [3]. Главной причиной введения единиц «бит» и «байт» явилась потребность определения объема запоминающих устройств, количества памяти, используемой компьютерной программой.

В ЭВМ каждый вводимый в машину символ кодируется словом длиной 8 бит. Для удобства введены более крупные, чем бит, единицы количества информации: 8 бит информации называют байтом. Байт – единица количества информации в Международной системе СИ. Значит, при вводе в ЭВМ одного символа, машине передается 1 байт информации. Для измерения информации используются и более крупные единицы:

$$1 \text{ Кбайт (килобайт)} = 2^{10} \text{ байт} = 1024 \text{ байта}$$

$$1 \text{ Мбайт (мегабайт)} = 2^{10} \text{ Кбайт} = 1024 \text{ Кбайта.}$$

В решении определения размера информации существуют такие подходы, как содержательный (вероятностный) и символьный (алфавитный) [4].

В содержательном подходе размер информации определяется объемом знаний, который несет сообщение получающему его человеку, то есть сообщение содержит информацию для человека, если заключенные в нем сведения являются для этого человека новыми и понятными, пополняющими его знания. Такой подход также часто называют субъективным, так как разные люди по-разному оценивают информацию об одном и том же предмете.

Алфавитный подход к измерению информации позволяет определить размер информации, заключенной в тексте, и является объективным, так как он не зависит от субъекта, воспринимающего текст.

Множество символов, используемых при записи текста, называется алфавитом. Полное количество символов в алфавите называется мощностью (размером) алфавита. Если допустить, что все символы алфавита встречаются в тексте с одинаковой частотой (равновероятно), то количество информации i , Кбайт, которое несет каждый символ, вычисляется по формуле:

$$i = \log_2 N,$$

где N – мощность алфавита.

Соответственно мощность алфавита N можно вычислить по формуле

$$N = 2^i.$$

Если весь текст состоит из K символов, то при алфавитном подходе размер содержащейся в нем информации I , Кбайт рассчитывается по формуле:

$$I = K \cdot i,$$

где i – информационный вес одного символа в используемом алфавите.

Использование алфавитного подхода позволит рассчитать размер информации, содержащейся в стандартах, и с помощью полученных данных станет возможным оценить полученный объем данных.

Данный подход был использован для стандартов раздела 77 – Metallургия; группы 77.120 – Цветные металлы; подгруппы 77.120.01 – Цветные металлы в целом. При этом ввели следующие термины:

Официальная часть: информация, включающая название стандарта, коды, дату введения, разработчиков и другие сведения до введения;

Содержательная часть: информация, включающая именно содержание стандарта, начиная с введения;

Общая информация: сумма официальной и содержательной частей.

Чтобы узнать, какой размер информации содержат стандарты раздела 77.120.01, данные стандарта необходимо представить в виде, удобном для последующей обработки пользователем, либо компьютером. Во внешнем представлении (для пользователя) все данные хранятся в виде файлов. Простейшими способами внешнего представления данных в стандарте являются: числовые данные, текст (последовательность символов), изображения (графики, фотографии, рисунки, схемы) [5, 6].

Таким образом, размер информации в стандартах складывается из размера числовых, текстовых и графических данных.

Рассмотрим пример расчета размера информации на примере стандарта ГОСТ 21073.1–75 «Металлы цветные. Определение величины зерна методом сравнения со шкалой микроструктур».

Так как размер числовых и текстовых данных стандарта зависит от использования выбранной кодировки, необходимо выбрать одну кодировку, в которой будем обрабатывать содержащуюся в нем информацию. Электронный запрос кода страницы, содержащей данный стандарт выдал кодировку стандарта UNICODE (UTF-8).

Стандарт содержит 6290 символов, с учетом русских и латинских букв, цифр и чисел и остальных знаков. В выбранной кодировке суммарный размер числовой и текстовой информации составил 11,21 Кбайт.

Кодирование графических изображений в стандарте осуществляется с помощью формата PNG (Portable Network Graphic) – растровый формат, используемый для размещения графики в Интернете. Всего данный стандарт содержит 31 изображение, суммарный размер графической информации составил 984 Кбайт.

Общий размер стандарта с учетом символьной, текстовой и графической информации составляет

$$I = 11,0 + 984 = 995,0 \text{ Кбайт.}$$

Используя данный пример расчета, были определены размеры информации остальных стандартов, включенных в подгруппу 77.120.01 Цветные металлы в целом [7], которые представлены в таблице.

Таблица

Размер информации в стандартах раздела ОКС 77.120.01 Цветные металлы в целом

| Временной интервал, гг. | Обозначение | Размер, Кбайт | | Размер по группам, Кбайт | | Число страниц, шт. | |
|-------------------------|-------------------|---------------|-----------------|---------------------------------|---------|--------------------|---------|
| | | офиц. части | содержат. части | общей информ. отдел. стандартов | суммар. | отдел. стандарта | суммар. |
| 1970-1979 | ГОСТ 21073.0-75 | 1,0 | 26,5 | 27,5 | 1046,5 | 18 | 32 |
| | ГОСТ 21073.1-75 | 1,0 | 994,0 | 995,0 | | 6 | |
| | ГОСТ 21073.2-75 | 1,0 | 7,0 | 8,0 | | 3 | |
| | ГОСТ 21073.3-75 | 1,0 | 5,0 | 6,0 | | 2 | |
| | ГОСТ 21073.4-75 | 1,0 | 9,0 | 10,0 | | 3 | |
| 1980-1989 | ГОСТ 24231-80 | 2,0 | 21,0 | 23,0 | 176,0 | 5 | 26 |
| | ГОСТ 28053-89 | 2,0 | 49,0 | 51,0 | | 6 | |
| | ГОСТ 28192-89 | 6,0 | 96,0 | 102,0 | | 15 | |
| 1990-1999 | ГОСТ 28873-90 | 10,0 | 20,0 | 30,0 | 30,0 | 15 | 15 |
| 2000-2009 | ГОСТ Р 53198-2008 | 12,0 | 97,0 | 109,0 | 109,0 | 23 | 23 |
| 2010-2019 | ГОСТ Р 54564-2011 | 11,0 | 234,0 | 245,0 | 320,0 | 62 | 78 |
| | ГОСТ 25086-2011 | 12,0 | 63,0 | 75,0 | | 16 | |

Рассмотрим, какой размер информации и какое количество действующих стандартов соответствует разным временным интервалам принятия этих стандартов.

На рис. представлено распределение размера информации, находящейся в стандартах, и количества стандартов в зависимости от временного интервала даты их принятия.

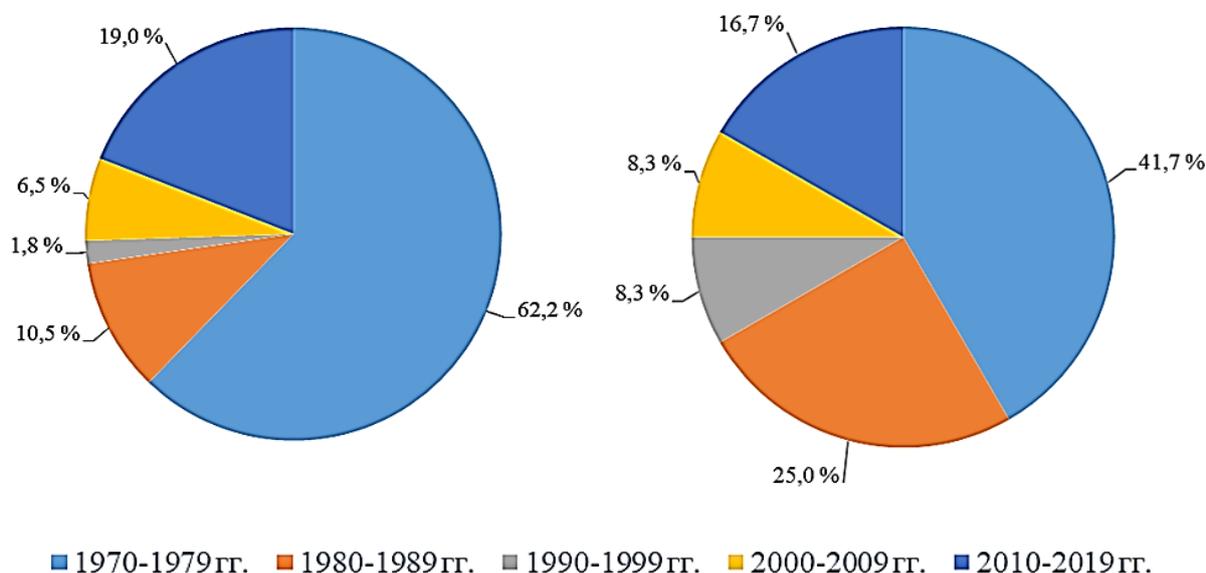


Рис. Сравнительный анализ размера информации и количества стандартов в зависимости от периода их принятия

Наибольший суммарный размер имеют стандарты, принятые в 1970-1979 годы, их доля составляет 62,2 % от общего размера, по числу стандартов эта группа также имеет наибольшую долю – 41,7 %. Другая ситуация со стандартами, принятыми в 1980-1989 гг. и 2010-2019 гг. Группа стандартов, принятых в 2010-2019 гг. имеет большую долю по размеру количества информации, при этом меньшее число стандартов, по сравнению со стандартами, принятыми в 1990-1999 гг.

Таким образом, с помощью алфавитного подхода измерения информации возможно объективно определить размер информации, заключенной в полном тексте стандарта. Расчет размера информации в стандартах раздела 77.120.02 «Цветные металлы в целом» показал, что наибольший размер имеют стандарты, принятые в 1970-1979 годы. Можно сказать, что по оценке количества стандартов складывается понимание о размере имеющейся в них информации, но эти сведения не дадут полного представления. Имеет смысл оценивать размер информации, например, в зависимости от соотношения различных форм представления данных.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Аджемов, А. С. Измерение количества информации / А. С. Аджемов, Б. П. Хромой // Мобильный бизнес: перспективы развития и реализации систем радиосвязи в России и за рубежом : сборник материалов XXXIII международной конференции РАЕН, Республика Индонезия, Лангкави, 15–17 апреля 2013 года. – Республика Индонезия, Лангкави: ЗАО «Национальный институт радио и инфокоммуникационных технологий», 2013. – С. 42-49.
2. Kazantseva, T. V. The Necessity and Complexity of Standardization / T. V. Kazantseva, N. K. Kazantseva, G. A. Tkachuk, A. L. Nevolina, N. A. Baganov, V. A. Aleksandrov, V. S. Kukhar // WSEAS TRANSACTIONS and CONTROL. – 2020. – Vol. 15. – P. 568-575.
3. ГОСТ 8.417-2002. ГСИ. Единицы величин : дата введения 2003–09–01 / Постановлением Государственного комитета Российской Федерации по стандартизации и метрологии // Техэксперт : Электронный фонд правовой и нормативно–технической документации. – URL: <https://docs.cntd.ru/document/1200031406?section=text>.
4. Руденко, Н. Б. Информатика: Понятие информации. Общая характеристика процессов сбора, передачи, обработки и накопления информации. Количество информации : методические указания / Н. Б. Руденко, Н. Н. Грачева, В. Н. Литвинов. – зерноград : Азово-Черноморский инженерный институт ФГБОУ ВПО «Донской государственный аграрный университет», 2015. – 58 с.
5. Волк, В. К. Информатика : учебное пособие для вузов / В. К. Волк. — Москва : Издательство Юрайт, 2021. — 207 с.
6. Пушкарёва, Т. П. Основы компьютерной обработки информации : учебное пособие / Т. П. Пушкарёва. – Красноярск : Сибирский федеральный университет (СФУ), 2016. – 180 с.
7. Техэксперт [Электронный ресурс] // Электронный фонд правовой и нормативно-технической документации. – 2021. – Режим доступа : <https://docs.cntd.ru>.

Светлана Комарова, Юлия Миндиярова, Виктория Серикова'

Svetlana Komarova, Yulia Mindiyarova, Victoria Serikova

**СТАНДАРТИЗАЦИЯ МЕТОДОВ ЗАЩИТЫ ОТ КОРРОЗИИ В
НЕФТЕПЕРЕРАБАТЫВАЮЩЕМ КОМПЛЕКСЕ**

**STANDARDIZATION OF METHODS OF CORROSION PROTECTION IN THE
OIL REFINING COMPLEX**

Российский химико-технологический университет имени Д. И. Менделеева

Mendeleev University of Chemical Technology

В настоящее время многие предприятия несут потери из-за выхода из строя оборудования, возникающего в результате коррозии. Причиной коррозии является не только

' Комарова С. – канд. техн. наук, доц.
Миндиярова Ю. – молодой специалист
Серикова В. – студент магистратуры