

Для оценки степени адекватности, реализующей спецификацию (1) – (3) математической модели анализируемому объекту, может быть использован широкий спектр показателей, разработанных в рамках анализа данных критериев. Это, в частности, коэффициент множественной детерминации; критерии Фишера, Стьюдена,

Хикса, Дарбина-Уотсона; средние относительные ошибки аппроксимации и прогноза, критерии смещения, согласованности поведения, информативности и т. д. Подробные смысловые характеристики и расчетные формулы критериев представлены, например в [1].

### Литература

1. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. Иркутск: Облформпечать, 1996. 320 с.

### References

1. Noskov S.I. Modeling technology for objects of unstable performance and data uncertainty. Irkutsk: Oblinformpechat', 1996. 320 s.

УДК 510.6

## Инновационные технологии обработки текстовых документов

Е.Г. Дулепов<sup>1\*</sup>, А.Н. Ефремова<sup>1</sup>

<sup>1</sup>Братский государственный университет, Макаренко 40, Братск, Россия  
Статья поступила 29.12.2011, принята 17.02.2012

*В статье предлагается алгоритм компьютерной обработки текстовых документов, ядром которого является разработанный универсальный алгоритм доказательства теорем логики высказываний. К основным текстовым документам авторы относят законы, договоры, контракты, декларации, соблюдение или выполнение которых необходимо постоянно контролировать, при этом полноценному компьютерному анализу и мониторингу текстовых документов должна предшествовать довольно сложная процедура формализации, связанная с трудностями, порождаемыми неоднозначностью семантики и синтаксиса естественных языков. Параметрами текстовых документов и предложений, входящих в них, являются смысл и логическая структура, имеющая истинностную оценку. Логической можно считать структуру документа, образованную истиной и ложью входящих в него предложений. Очевидно, что смысл любого текстового документа ясен его создателям. Отсюда следует, что для компьютерной обработки документа необходимо сохранить его логическую структуру и удалить смысл. Такая процедура, как известно, называется формализацией и предполагает замену каждого предложения документа буквой, а логических связей между предложениями (буквами) – символами известных логических операций языка вида: НЕ, И, ИЛИ, ЕСЛИ – ТО. В результате формализации задание текстового документа будет представлено решением логического уравнения  $f(x_1, \dots, x_n) = 1$ , где  $f(x_1, \dots, x_n)$  – функция алгебры логики. Выполняя миллионы операций в секунду, компьютер существенно упростит по форме это решение. При необходимости, после компьютерной обработки можно получить новый для пользователя, но уже более удобный текстовый документ, адекватный исходному. Эта процедура называется интерпретацией. В статье предлагаются анализ и мониторинг текстового документа путем компьютерного доказательства существования теоремы вида  $(f_{\text{ГИПОТЕЗА}} \cdot f(x_1, \dots, x_n = 1) \Rightarrow f_{\text{СЛЕДСТВИЕ}}$ . Условия, или гипотезы, воздействующие на документ, определяют внешние причины, например, рынок. Верификация алгоритма обработки текстовых документов демонстрируется решениями трех задач: мониторинга коммерческого договора, упрощения логического выражения и оценки истинности сложной конструктивной дилеммы, при этом формализация выполняется только для первой задачи.*

**Ключевые слова:** текстовые документы, теоремы, гипотезы, формализация, интерпретация, верификация.

## Innovation techniques for text documents processing

E.G. Dulepov<sup>1\*</sup>, A.N. Efremova<sup>1</sup>

<sup>1</sup>Bratsk State University, 40, Makarenko str., Bratsk, Russia  
Received 29.12.2011; Accepted 17.02.2012

*In the article the authors propose the algorithm for electronic text documents processing, its core being the universal algorithm of the assertion-level logic theorem proving developed by the authors and published in № 2 (10), 2011 «Systems. Methods. Technologies» scientific journal. According to the authors, the main text documents to be observed and constantly monitor are laws, agreements, contracts, declarations. The parameters of text documents are the meaning and logical structure having a truth-value assessment. The meaning of any text document is usually clear to its authors and plays no role later on, because to electronically process the document, it is proposed to preserve only its logical structure and remove its meaning. This procedure is known as formalization and involves the*

\* E-mail address: iipm@brstu.ru

replacement of every sentence in the document by a letter, and the replacement of logical connectives between sentences by symbols of the known language logical operations. As a result of formalization, the task of the text document will be represented by the solution of a logical equation concerning the function of algebra of logic representing the document. A computer can considerably simplify the solution and, if necessary, offer it in an easier-to-use form. The analysis and monitoring of the text document is proposed to be done by means of a computer proving of the existence of a specific type theorem taking into consideration the influence of the environment. The algorithm operation is demonstrated in the article with the help of the following problems solving: commercial contract monitoring, logical expression reduction and the truth assessment of a complex constructive dilemma.

**Keywords:** text documents, theorems, hypotheses, formalization, interpretation.

Полноценный анализ и мониторинг текстовых документов может выполнить только компьютер. К сожалению, любой язык, особенно русский, из-за своего богатого содержания имеет неоднозначные синтаксис и семантику, непонятные компьютеру, в результате чего компьютерная обработка таких текстовых документов будет невозможна.

К основным текстовым документам относят, например, законы, договоры, контракты, декларации, соблюдение или выполнение которых надо постоянно контролировать.

Параметрами текстовых документов и предложений, входящих в них, являются смысл и логическая структура, имеющая истинностную оценку. Логической называют структуру, образованную истиной и ложью входящих в нее предложений.

Смысл создателям любого документа обычно ясен и в дальнейшей судьбе документа никакой роли не играет, поэтому для компьютерной обработки документа необходимо в нем сохранить логическую структуру и удалить смысл. Такая процедура называется **формализацией**. Формализация предполагает замену каждого предложения документа буквой и замену логических связей между предложениями (буквами) символами известных логических операций языка вида: НЕ, И, ИЛИ, ЕСЛИ – ТО. Тогда текстовый документ превратиться в **формулу** функции алгебры логики  $f(x_1, \dots, x_n)$ .

Выполняя миллионы операций в секунду, компьютер существенно упростит эту формулу. При необходимости после компьютерной обработки можно получить для пользователя новый, но уже более удобный текстовый документ, адекватный исходному. Эта процедура называется **интерпретацией**.

Анализ и мониторинг документа могут быть выполнены путем компьютерного решения логического урав-

нения  $f(x_1, \dots, x_n) = 1$  или путем компьютерного доказательства существования теоремы вида

$$(f_{\text{ГИПОТЕЗА}}, f(x_1, \dots, x_n = 1) \Rightarrow f_{\text{СЛЕДСТВИЕ}})$$

Универсальный алгоритм доказательства теорем и его программная реализация частично опубликованы в [1, 2]. Условия, или гипотезы, воздействующие на документ, определяют внешние причины, например, рынок.

Приведем несколько примеров компьютерной обработки формализованных текстовых документов и сделаем важные выводы.

**Пример 1.** Компьютерная обработка коммерческого договора: «Если курс доллара высок ( $X$ ), то товар следует продавать в Китае ( $Y$ ), в противном случае ( $\bar{X}$ ), товар следует продавать в Братске» ( $Z$ ).

Смысл договора очевиден, после формализации договор будет представлен логическим уравнением или просто задан числом 202 (формализация):

$$f(x, y, z) = w_{202} = k(7,6,3,1) = m(5,4,2,0) = xy \vee yz \vee \bar{x}z = xy \vee \bar{x}z = 1$$

Сокр ДНФ                      Мин ДНФ

Существуют компьютерные программы построения сокращенных и минимальных дизъюнктивных нормальных форм (ДНФ-функций).

Факт выполнения договора в условиях рынка означает существование теоремы:

$$f_{\text{ГИПОТЕЗА}}, f(x, y, z) = 1 \Rightarrow f_{\text{СЛЕДСТВИЕ}}$$

Компьютерные решения этой задачи для случаев  $x = 1$  приведены на рис. 1 (в первой строке решений показана запись договора на языке программирования).

$$f_{\text{оп}} = x \cdot (x \cdot y \vee \neg x \cdot z)$$

$$f_{\text{зак}} = z$$

$$f_{\text{оп}} \Rightarrow f_{\text{зак}}$$

$$x \cdot (x \cdot y \vee \neg x \cdot z) \Rightarrow z$$

	00	01	02	03	04	05	06	07
x	0	0	0	0	1	1	1	1
y	0	0	1	1	0	0	1	1
z	0	1	0	1	0	1	0	1
f <sub>оп</sub>	0	0	0	0	0	1	1	1
f <sub>зак</sub>	0	1	0	1	0	1	0	1
f <sub>оп</sub> → f <sub>зак</sub>	1	1	1	1	1	1	0	1

$(f_{\text{оп}} \neq 0) \cdot (f_{\text{оп}} \leq f_{\text{зак}}) = (E_{\text{оп}} \neq \emptyset) \cdot (E_{\text{оп}} \subseteq E_{\text{зак}}) = 0$   
 $f_{\text{оп}} \Rightarrow f_{\text{зак}}$  — нетеорема

Договор нарушен

$$f_{\text{оп}} = x \cdot (x \cdot y \vee \neg x \cdot z)$$

$$f_{\text{зак}} = y$$

$$f_{\text{оп}} \Rightarrow f_{\text{зак}}$$

$$x \cdot (x \cdot y \vee \neg x \cdot z) \Rightarrow y$$

	00	01	02	03	04	05	06	07
x	0	0	0	0	1	1	1	1
y	0	0	1	1	0	0	1	1
z	0	1	0	1	0	1	0	1
f <sub>оп</sub>	0	0	0	0	0	1	1	1
f <sub>зак</sub>	0	0	1	1	0	0	1	1
f <sub>оп</sub> → f <sub>зак</sub>	1	1	1	1	1	1	1	1

$(f_{\text{оп}} \neq 0) \cdot (f_{\text{оп}} \leq f_{\text{зак}}) = (E_{\text{оп}} \neq \emptyset) \cdot (E_{\text{оп}} \subseteq E_{\text{зак}}) = 1$   
 $f_{\text{оп}} \Rightarrow f_{\text{зак}}$  — теорема

Договор выполняется

Рис. 1. Компьютерный мониторинг договора ( $x, xy \vee \bar{x}z = 1; x = 1$ ).

$$f_{он} = (x \vee y) \cdot (x \vee z)$$

$$f_{зак} = (x \vee y) \cdot (\neg y \vee z) \cdot (x \vee z)$$

$$f_{он} \Rightarrow f_{зак}$$

$$(x \vee y) \cdot (x \vee z) \Rightarrow (x \vee y) \cdot (\neg y \vee z) \cdot (x \vee z)$$

	00	01	02	03	04	05	06	07
x	0	0	0	0	1	1	1	1
y	0	0	1	1	0	0	1	1
z	0	1	0	1	0	1	0	1
f <sub>он</sub>	0	0	0	1	1	1	1	1
f <sub>зак</sub>	0	0	0	1	1	1	0	1
f <sub>он</sub> → f <sub>зак</sub>	1	1	1	1	1	1	0	1

$(f_{он} \neq 0) \cdot (f_{он} \leq f_{зак}) = (E_{он} \neq \emptyset) \cdot (E_{он} \subseteq E_{зак}) = 0$   
 $f_{он} \Rightarrow f_{зак}$  — **нетеорема**  
 Сомножитель  $(\bar{y} \vee z)$  – не лишний

$$f_{он} = (x \vee y) \cdot (\neg y \vee z)$$

$$f_{зак} = (x \vee y) \cdot (\neg y \vee z) \cdot (x \vee z)$$

$$f_{он} \Rightarrow f_{зак}$$

$$(x \vee y) \cdot (\neg y \vee z) \Rightarrow (x \vee y) \cdot (\neg y \vee z) \cdot (x \vee z)$$

	00	01	02	03	04	05	06	07
x	0	0	0	0	1	1	1	1
y	0	0	1	1	0	0	1	1
z	0	1	0	1	0	1	0	1
f <sub>он</sub>	0	0	0	1	1	1	0	1
f <sub>зак</sub>	0	0	0	1	1	1	0	1
f <sub>он</sub> → f <sub>зак</sub>	1	1	1	1	1	1	1	1

$(f_{он} \neq 0) \cdot (f_{он} \leq f_{зак}) = (E_{он} \neq \emptyset) \cdot (E_{он} \subseteq E_{зак}) = 1$   
 $f_{он} \Rightarrow f_{зак}$  — **теорема**  
 Сомножитель  $(x \vee z)$  – лишний

Рис. 2. Компьютерное упрощение выражения  $(x \vee y)(\bar{y} \vee z)(x \vee z)$ .

При мониторинге текстовых документов после их формализации и при синтезе цифровых схем может возникнуть необходимость компьютерного упрощения полученных выражений.

Существенно можно упростить, например, формулу функции алгебры логики, удаляя в произвольных логических произведениях лишние сомножители, не равные единице, и в произвольных логических суммах – лишние слагаемые, не равные нулю. Для решения этой задачи можно воспользоваться утверждениями из [2]:

1. Если  $a \vee b \vee c \Rightarrow a \vee c$  – теорема, то слагаемое  $b$  – лишнее;
2. Если  $a \cdot c \Rightarrow a \cdot b \cdot c$  – теорема, то сомножитель  $b$  – лишний.

**Пример 2.** Какой из сомножителей  $(\bar{y} \vee z)$  или  $(x \vee z)$  в логическом произведении  $(x \vee y)(\bar{y} \vee z)(x \vee z)$  является лишним?

Компьютерное решение на рис. 2:

**Пример 3.** Доказать истинность древнегреческой сложной конструктивной дилеммы:

$$s \rightarrow p, c \rightarrow d, s \vee c \Rightarrow p \vee d.$$

Компьютерное решение (дилемма является теоремой) на рис. 3.

Изложенный выше материал частично опубликован в [1, 2, 3] и обсуждался 17 ноября 2011 г. на научно-техническом семинаре в рамках деловой программы выставки «Братск: строительство, энергетика ЖКХ, газификация», организованной администрацией г. Братска и ООО «СибЭкспоСервис-Н».

*Литература*

1. Дулепов Е.Г., Ефремова А.Н. Структура и возможности универсального алгоритма доказательства теорем // Системы. Методы. Технологии. 2011. № 2 (10). С. 71-73.
2. Дулепов Е.Г. Логика как инструмент исследования знаний. Братск: Брат. гос. ун-т, 2011. 133 с. Деп. в ВИНТИ 27.01.11, № 26-B2011
3. Дулепов Е.Г. Теоретические основы вычислительной техники. Братск: Брат. гос. ун-т, 2008. 236 с. Деп. в ВИНТИ. 04.07.2008, № 578 - B2008.

$$s \rightarrow p, c \rightarrow d, s \vee c \Rightarrow p \vee d.$$

$$f_{он} = (s \rightarrow p) \cdot (c \rightarrow d) \cdot (s \vee c)$$

$$f_{зак} = (p \vee d)$$

$$f_{он} \Rightarrow f_{зак}$$

$$(s \rightarrow p) \cdot (c \rightarrow d) \cdot (s \vee c) \Rightarrow (p \vee d)$$

	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
c	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
d	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
p	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
s	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
f <sub>он</sub>	0	0	0	1	0	0	1	0	0	0	0	1	0	1	1	
f <sub>зак</sub>	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1	
f <sub>он</sub> → f <sub>зак</sub>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	

$(f_{он} \neq 0) \cdot (f_{он} \leq f_{зак}) = (E_{он} \neq \emptyset) \cdot (E_{он} \subseteq E_{зак}) = 1$   
 $f_{он} \Rightarrow f_{зак}$  — **теорема**

Рис. 3. Компьютерное доказательство дилеммы как теоремы.

**Выводы**

1. Основным инструментом исследования различных форм знания является современная конструктивная логика.
2. К сожалению, в настоящее время логика как дисциплина практически исключена из программ средних и высших учебных заведений России, и этим могут воспользоваться мошенники, чиновники-взяточники и работодатели, нарушающие трудовые соглашения.
3. Необходима модернизация программ изучения русского языка, поскольку сейчас в школах и вузах из них практически исключена рациональная часть языка, которая отвечает за построение логических структур текстов и правильное решение вопроса об их истинности или ложности.

*References*

1. Dulepov E.G., Efremova A.N. The structure and resource of the universal theorem proving algorithm // Sistemy. Metody. Tekhnologii. 2011. № 2(10). S. 71- 73.
2. Dulepov E.G. Logic as an instrument of knowledge analysis. Bratsk: Brat. gos. un-t, 2011. 133 s. Dep. v VINITI 27.01.11 №26-B2011.
3. Dulepov E.G. Computer engineering theory. Bratsk: Brat. gos. un-t, 2008. 236 s. Dep. v VINITI. 04.07.2008, № 578 - B2008.