

УДК 002.66

UDC 002.66

**ПОДХОДЫ К ОПЕРАТИВНОЙ ИДЕНТИФИКАЦИИ ФОРМАЛИЗОВАННЫХ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ В АВТОМАТИЗИРОВАННЫХ ДЕЛОПРОИЗВОДСТВАХ**

**APPROACHES TO THE RAPID IDENTIFICATION OF FORMALIZED ELECTRONIC DOCUMENTS IN AUTOMATED RECORDS MANAGEMENT**

Королев Игорь Дмитриевич  
д.т.н., профессор

Korolyov Igor Dmitrievich  
Dr.Sci.Tech., professor

Носенко Сергей Владимирович  
*Филиал Военной академии связи  
(г. Краснодар), Краснодар, Россия*

Nosenko Sergey Vladimirovich  
*Krasnodar branch of Military Academy of communication, Krasnodar, Russia*

В данной статье представлен подход к представлению различных видов электронных документов позволяющий создать унифицированный информационно-поисковый тезаурус

The article presents different types of electronic documents allowing to create the unique searching-for-information thesaurus

Ключевые слова: АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ ДЕЛОПРОИЗВОДСТВА, ЗОНЫ ДОКУМЕНТА, ПРЕДИКАТА УЗНАВАНИЯ ВИДА ДОКУМЕНТА

Keywords: AUTOMATION SYSTEMS, ZONE OF DOCUMENT, PREDICATE RECOGNITION OF TYPE OF DOCUMENT

При переходе на электронный документооборот становится возможным применение технологий автоматической обработки текстов. На первый план в электронном документообороте выходят: автоматическое концептуальное индексирование, применение автоматических поисковых запросов, автоматическое рубрицирование и аннотирование отдельных документов их кластеризация. Актуальность работы состоит в том, что предложенный авторами способ позволяет проводить все вышеперечисленные операции с наименьшими временными затратами.

Формализованные документы, используемые в официальной переписке и обеспечении внутренней работы организаций, имеют определенную структуру, соответствующую виду документа и исполняются служебно-деловым стилем с соблюдением правил русской орфографии и пунктуации, обеспечивают точное и однозначное восприятие изложенной в нем информации. Применяемые термины должны соответствовать принятой терминологии и употребляться в одном и том же значении.

Поскольку язык для составления документов и сопутствующая деятельность, связанная с обеспечением функционирования делопроизводства

лежит в области интеллектуальной деятельности человека, анализ сложности моделирования всевозможных видов документов, связанных с используемым в документах естественным языком привели к следующим выводам [1]:

1. Формальный язык документов, как и человеческий язык, явление дискретное, естественно, они должны описываться средствами дискретной математики. Причем выбор средств указанного типа ограничен: языки программирования, логические исчисления, языки теории алгоритмов, аппарат теории графов.

2. При анализе применения языков программирования или языков теории алгоритмов пришлось столкнуться со сложностью описания алгоритмов, то есть процедур с однозначным исходом по причине многозначности человеческого языка. Языки программирования и теории алгоритмов – это такие языки, которые могут описывать только однозначные функции.

3. Логические исчисления, а именно: исчисления высказываний и исчисления предикатов не имеют четкой и полной алгебраической системы. Это сделано только в исчислении высказываний. В результате мы имеем алгебру логики и аппарат булевых уравнений, который вызывает неудобство, заключающееся в том, что в алгебре логики используются лишь двоичные знаки, в то время как в естественном языке фигурируют буквенные, то есть многозначные символы.

4. Попытка устранения данного недостатка – обращение к аппарату многозначной логики, но многозначная логика развита только в описании однозначных функций, а не отношений. Развитие же в этом направлении многозначной логики принудительно приводит к алгебре конечных предикатов. Чтобы иметь возможность записывать самые общие уравнения многозначной логики, в правой их части нет необходимости ставить произвольные формулы, достаточно писать константы. Необязательно использовать все константы, достаточно взять всего два знака: 0 и 1. Но как только

мы так поступим, немедленно приходим к понятию конечного предиката, а, следовательно, и к алгебре конечных предикатов.

5. Использование исчисления предикатов для целей математического описания человеческого языка также наталкивается на определенную трудность: исчисление очень слабо развито применительно к нуждам описания конечных объектов. Исчисление предикатов не располагает даже средствами для формульной записи любых индивидуальных конечных отношений. Вместе с тем, человеческий язык – явление сугубо конечное и он требует для своей формализации аппарата конечной математики. Пытаясь алгебраизировать конечный фрагмент исчисления предикатов, мы не сможем прийти ни к чему иному, как только к алгебре конечных предикатов.

6. Обратившись к аппарату теории графов, мы обнаружим, что, хотя он и используется для описания конечных отношений, однако совершенно не содержит в себе выразительных средств для записи этих отношений в виде уравнений некоторой алгебры. Если же мы захотим перевести информацию, содержащуюся в графах, на язык таблиц, то увидим, что с помощью графов выражаются именно конечные предикаты.

Таким образом, какой бы путь мы ни избрали при разработке приемлемых формальных средств для математического описания человеческого языка, мы неизбежно приходим к алгебре конечных предикатов. Вместе с тем установлено, что алгебра конечных предикатов полна [2], то есть на ее языке могут быть описаны любые конечные отношения. Поэтому любой другой математический аппарат, предназначенный для описания произвольных конечных отношений, в логическом смысле обязательно будет равносителен алгебре конечных предикатов.

Важнейшим вопросом для обеспечения оперативной работы автоматизированной системы делопроизводства на первом этапе обработки поступающего документа является извлечение метаданных, позволяющих максимально повысить оперативность всех последующих внутренних про-

цессов обработки документов в автоматизированном делопроизводстве: для формализованного представления данных (учет), при выполнении запросов к данным, при работе механизмов обработки запросов.

С целью построения такой модели извлечения метаданных из документов необходимо определиться со всем разнообразием видов документов, определения их зон и списков метаданных, относящихся к той или иной зоне документа для их последующего выделения. Количество различных зон документа может расширяться, но так как количество видов формализованных документов, поступающих в систему конечно, соответственно количество типов зон документов конечно и много меньше количества документов. Примерами зон могут быть: заголовки, текст, списки и графика как содержимое текстов, обращения, зоны согласования и утверждения.

Используем следующий порядок определения вида документа и его зон. На рисунке 1 представлен пример документа, математическое описание вида которого в общем виде выразится конечным предикатом  $v_j(Z, L)$ , где  $V = \{v_1, v_2, \dots, v_m\}$  – множество видов документа,  $j = \{1, 2, \dots, m\}$ ;  $m$  – количество всех используемых видов документов,  $Z = \{z_1, z_2, \dots, z_n\}$  – множество конечных предикатов известных зон документа,  $n$  – количество всех зон документов,  $L = \{l_1, l_2, \dots, l_q\}$  – множество конечных предикатов узнавания ключевых слов,  $q$  – количество всех используемых ключевых слов.

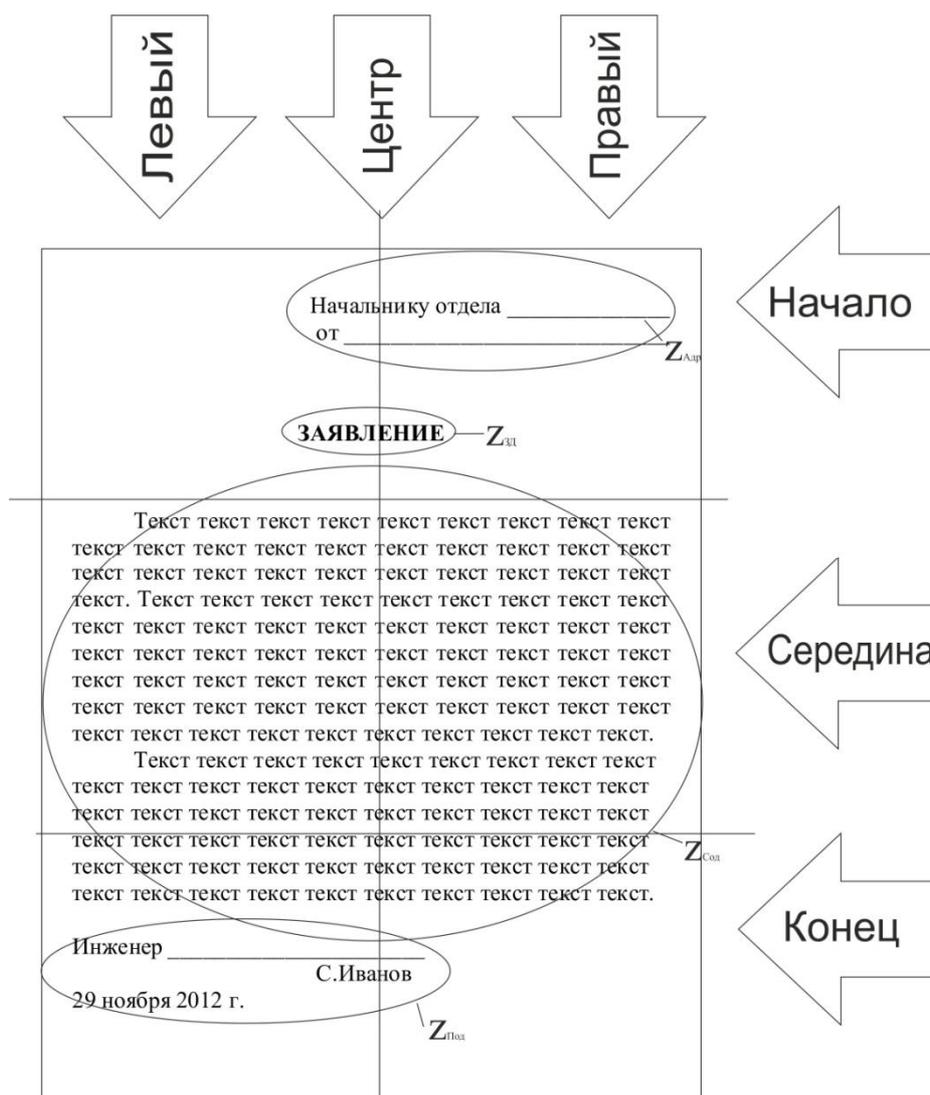


Рис. 1 – Пример документа, поступающего в автоматизированное делопроизводство.

В представленном примере документа для выделения зон документа используется характеристики содержимого, имеющего единую структуру. Например, в качестве зоны принимается содержимое одного (или большего количества) абзаца однотипно выделенного (курсив, подчеркивание и т.д.), или зона представляет собой содержимое строк, выровненных, например, по краю (правому, левому) или по центру.

Каждую зону возможно представить в виде конечного предиката  $z_i(T, L)$ , где  $T$  – множество конечных предикатов узнавания характеристик тек-

ста  $t$ ,  $L=\{l_1, l_2, \dots, l_q\}$  – множество конечных предикатов узнавания ключевых слов,  $q$  – количество всех используемых ключевых слов.

Используя в дальнейшем алгебру конечных предикатов, возможно описание и построение математической модели определения вида документа по характеристикам текста.

Рассмотрим вышеописанные выражения применительно к некоторым видам документов: Приказ, Постановление, Регламент, Положение, Инструкция, Донесение, Заявление, Договор, Контракт, План, Отчет, Перечень, Протокол, Акт, Справка, Служебное письмо, Заявка.

Введем множество переменных  $t_1, t_2, t_3, t_4$  – множество характеристик текста – 4 переменные с соответствующими алфавитами величиной 3, 4, 4, 3 (Таб. 1). Показателями характеристики текста необходимыми при определении зоны текста являются:

$t_1$  - месторасположение в документе: начало, середина, конец;

$t_2$  - месторасположение на строке: слева, по центру, справа, по ширине;

$t_3$  - выделение текста: нет, жирным, курсивом, подчеркивание;

$t_4$  - размерность шрифта (можно задать промежутками): 0-10, 11-14, 15-∞.

Таблица 1 - Положение и характеристики зон в документах (пример)

Характеристики текста Зона документа $Z=\{z_1, z_2, \dots, z_n\}$	$t_1$ - месторасположение в документе			$t_2$ - месторасположение на строке				$t_3$ - выделение текста				$t_4$ - размерность шрифта		
	начало	середина	конец	По левому	по центру	По правому	По ширине	нет	жирным	курсивом	подчеркивание	0-10	11-14	15-∞
Переменные	1	2	3	1	2	3	4	1	2	3	4	1	2	3
Угловой штамп	+			+				+	+			+		

Характеристики текста  Зона доку- мента $Z=\{z_1, z_2, \dots, z_n\}$	t <sub>1</sub> - месторасположение в документе			t <sub>2</sub> - месторасположение на строке				t <sub>3</sub> - выделение текста				t <sub>4</sub> - размерность шрифта		
	начало	середина	конец	По левому	по центру	По правому	По ширине	нет	жирным	курсивом	подчеркивание	0-10	11-14	15-∞
Переменные	1	2	3	1	2	3	4	1	2	3	4	1	2	3
Согласование	+		+	+	+	+		+					+	
Утверждение	+					+		+					+	
Заголовок документа	+				+			+	+				+	+
Адресат	+				+	+		+	+		+		+	
Содержание		+		+			+	+	+	+	+		+	
Подпись документа			+	+	+			+	+				+	
Отметка об исполнителе			+	+				+					+	

Правило построения предиката узнавания зоны документа по характеристикам текста, выразиться следующей формулой:

$$z_i(T) = \bigwedge_{h=1}^m \bigvee_{a=1}^n t_h^a,$$

где  $t_h^a$  – предикат узнавания значения  $ah$ -той переменной текста;  $m$  – количество переменных текста,  $n$  – величина алфавита  $h$ -той переменной текста.

Выражение предикатов зон документов через переменные текста  $t_1, t_2, t_3, t_4$ :

$$z_{\text{Уш}}^T = t_1^1 t_2^1 (t_3^1 \vee t_3^2) t_4^1;$$

$$z_{\text{Сог}}^T = (t_1^1 \vee t_1^2) (t_2^1 \vee t_2^2 \vee t_2^3) t_3^1 t_4^2;$$

$$z_{\text{Ут}}^T = t_1^1 t_2^3 t_3^1 t_4^2;$$

$$z_{\text{Зд}}^T = t_1^1 t_2^2 (t_3^1 \vee t_3^2) (t_4^2 \vee t_4^3);$$

$$z_{\text{Адр}}^T = t_1^1 (t_2^2 \vee t_2^3) (t_3^1 \vee t_3^2 \vee t_3^3) t_4^2;$$

$$z_{\text{Сод}}^T = t_1^2 (t_2^1 \vee t_2^4) (t_3^1 \vee t_3^2 \vee t_3^3 \vee t_3^4) t_4^2;$$

$$z_{\text{Под}}^T = t_1^3(t_2^1 \vee t_2^2)(t_3^1 \vee t_3^2)t_4^2;$$

$$z_{\text{Исп}}^T = t_1^3 t_2^1 t_3^1 t_4^2;$$

Внешне предикаты зон документа различны, но необходима проверка их уникальность, т.е. однозначность определения зоны в документе, что позволяет нам сделать алгебра конечных предикатов путем приведения наших предикатов к совершенной дизъюнктивной нормальной форме (далее – СДНФ) и проведем анализ используемых в них конституэнт единицы.

$$z_{\text{УШ}}^T = t_1^1 t_2^1 (t_3^1 \vee t_3^2) t_4^1 = t_1^1 t_2^1 t_3^1 t_4^1 \vee t_1^1 t_2^1 t_3^2 t_4^1;$$

$$\begin{aligned} z_{\text{Сор}}^T &= (t_1^1 \vee t_1^3)(t_2^1 \vee t_2^2 \vee t_2^3) t_3^1 t_4^2 = (t_1^1 \vee t_1^3)(t_2^1 t_3^1 t_4^2 \vee t_2^2 t_3^1 t_4^2 \vee t_2^3 t_3^1 t_4^2) \\ &= t_1^1 t_2^1 t_3^1 t_4^2 \vee t_1^1 t_2^2 t_3^1 t_4^2 \vee t_1^1 t_2^3 t_3^1 t_4^2 \vee t_1^3 t_2^1 t_3^1 t_4^2 \vee t_1^3 t_2^2 t_3^1 t_4^2 \vee t_1^3 t_2^3 t_3^1 t_4^2; \end{aligned}$$

$$z_{\text{УТ}}^T = t_1^1 t_2^3 t_3^1 t_4^2;$$

$$\begin{aligned} z_{\text{ЗД}}^T &= t_1^1 t_2^2 (t_3^1 \vee t_3^2)(t_4^2 \vee t_4^3) = (t_1^1 t_2^2 t_3^1 \vee t_1^1 t_2^2 t_3^2)(t_4^2 \vee t_4^3) \\ &= t_1^1 t_2^2 t_3^1 t_4^2 \vee t_1^1 t_2^2 t_3^2 t_4^2 \vee t_1^1 t_2^2 t_3^1 t_4^3 \vee t_1^1 t_2^2 t_3^2 t_4^3; \end{aligned}$$

$$\begin{aligned} z_{\text{Адр}}^T &= t_1^1 (t_2^2 \vee t_2^3)(t_3^1 \vee t_3^2 \vee t_3^3) t_4^2 = (t_1^1 t_2^2 \vee t_1^1 t_2^3)(t_3^1 t_4^2 \vee t_3^2 t_4^2 \vee t_3^3 t_4^2) \\ &= t_1^1 t_2^2 t_3^1 t_4^2 \vee t_1^1 t_2^2 t_3^2 t_4^2 \vee t_1^1 t_2^2 t_3^3 t_4^2 \vee t_1^1 t_2^3 t_3^1 t_4^2 \vee t_1^1 t_2^3 t_3^2 t_4^2 \vee t_1^1 t_2^3 t_3^3 t_4^2; \end{aligned}$$

$$\begin{aligned} z_{\text{Сод}}^T &= t_1^2 (t_2^1 \vee t_2^4)(t_3^1 \vee t_3^2 \vee t_3^3 \vee t_3^4) t_4^2 = (t_1^2 t_2^1 \vee t_1^2 t_2^4)(t_3^1 t_4^2 \vee t_3^2 t_4^2 \vee t_3^3 t_4^2 \vee t_3^4 t_4^2) \\ &= t_1^2 t_2^1 t_3^1 t_4^2 \vee t_1^2 t_2^1 t_3^2 t_4^2 \vee t_1^2 t_2^1 t_3^3 t_4^2 \vee t_1^2 t_2^1 t_3^4 t_4^2 \vee t_1^2 t_2^4 t_3^1 t_4^2 \vee t_1^2 t_2^4 t_3^2 t_4^2 \vee t_1^2 t_2^4 t_3^3 t_4^2 \vee t_1^2 t_2^4 t_3^4 t_4^2; \end{aligned}$$

$$\begin{aligned} z_{\text{Под}}^T &= t_1^3 (t_2^1 \vee t_2^2)(t_3^1 \vee t_3^2) t_4^2 = (t_1^3 t_2^1 \vee t_1^3 t_2^2)(t_3^1 t_4^2 \vee t_3^2 t_4^2) \\ &= t_1^3 t_2^1 t_3^1 t_4^2 \vee t_1^3 t_2^1 t_3^2 t_4^2 \vee t_1^3 t_2^2 t_3^1 t_4^2 \vee t_1^3 t_2^2 t_3^2 t_4^2; \end{aligned}$$

$$z_{\text{Исп}}^T = t_1^3 t_2^1 t_3^1 t_4^2.$$

Выявлены повторяющиеся конституэнты единицы:

$$t_1^1 t_2^3 t_3^1 t_4^2 \text{ в } z_{\text{Сор}}^T, z_{\text{УТ}}^T, z_{\text{Адр}}^T;$$

$$t_1^1 t_2^2 t_3^1 t_4^2 \text{ в } z_{\text{Сор}}^T, z_{\text{ЗД}}^T, z_{\text{Адр}}^T;$$

$$t_1^3 t_2^1 t_3^1 t_4^2 \text{ в } z_{\text{Сор}}^T, z_{\text{Под}}^T, z_{\text{Исп}}^T;$$

$$t_1^2 t_2^2 t_3^1 t_4^2 \text{ в } z_{\text{Сор}}^T, z_{\text{Под}}^T;$$

$$t_1^1 t_2^2 t_3^2 t_4^2 \text{ в } z_{\text{ЗД}}^T, z_{\text{Адр}}^T.$$

Из полученного анализа конституэнт единицы СДНФ предикатов видно, что однозначно с использованием только текстуальных характеристик определяется  $Z_{уш}$ ,  $Z_{Сод}$  (угловой штамп и содержание). Остальные предикаты однозначного определения зоны не дают.

Устранить неоднозначность можно, по крайней мере, двумя способами:

А) Организационно – большей формализацией используемых документов. В представленном случае выявить необходимые ограничения возможно путем удаления повторяющихся конституэнт единицы в предикатах, где их наличие необязательно, в итоге получится:

$$Z_{уш}^T = t_1^1 t_2^1 t_3^1 t_4^1 \vee t_1^1 t_2^1 t_3^2 t_4^1;$$

$$Z_{Сог}^T = t_1^1 t_2^1 t_3^1 t_4^2 \vee t_1^1 t_2^2 t_3^1 t_4^2 \vee t_1^3 t_2^3 t_3^1 t_4^2;$$

$$Z_{ур}^T = t_1^1 t_2^3 t_3^1 t_4^2;$$

$$Z_{Зд}^T = t_1^1 t_2^2 t_3^2 t_4^2 \vee t_1^1 t_2^2 t_3^1 t_4^2 \vee t_1^1 t_2^2 t_3^2 t_4^2;$$

$$Z_{Адп}^T = t_1^1 t_2^2 t_3^4 t_4^2 \vee t_1^1 t_2^3 t_3^2 t_4^2 \vee t_1^1 t_2^3 t_3^4 t_4^2;$$

$$Z_{Сод}^T = t_1^2 t_2^1 t_3^1 t_4^2 \vee t_1^2 t_2^1 t_3^2 t_4^2 \vee t_1^2 t_2^1 t_3^3 t_4^2 \vee t_1^2 t_2^4 t_3^1 t_4^2 \vee t_1^2 t_2^4 t_3^2 t_4^2 \vee t_1^2 t_2^4 t_3^3 t_4^2 \vee t_1^2 t_2^4 t_3^4 t_4^2;$$

$$Z_{Под}^T = t_1^3 t_2^1 t_3^2 t_4^2 \vee t_1^3 t_2^2 t_3^1 t_4^2 \vee t_1^3 t_2^2 t_3^2 t_4^2;$$

$$Z_{Исп}^T = t_1^3 t_2^1 t_3^1 t_4^2;$$

Соответственно новые требования к расположению и характеристикам зон в документах в отличие от Таблицы 1 для однозначного их определения отражены в таблице 2:

Таблица 2 - Новые требования к расположению и характеристикам зон в документах

Характеристики текста  Зона документа $Z=\{z_1, z_2, \dots, z_n\}$	t <sub>1</sub> - месторасположение в документе			t <sub>2</sub> - месторасположение на строке				t <sub>3</sub> - выделение текста				t <sub>4</sub> - размерность шрифта		
	начало	середина	конец	По левому	по центру	По правому	По ширине	нет	жирным	курсивом	подчеркиванием	0-10	11-14	15-∞
Переменные	1	2	3	1	2	3	4	1	2	3	4	1	2	3
Угловой штамп	+			+				+	+			+		
Согласование	+			+	+			+					+	
Утверждение	+					+		+					+	
Заголовок документа	+				+			+	+				+	+
Адресат	+				+	+			+		+		+	
Содержание		+		+			+	+	+	+	+		+	
Подпись документа			+	+	+			+	+				+	
Отметка об исполнителе			+	+				+					+	

т.е. необходимо дополнительно:

запретить располагать в конце документа и справа на строке согласование;

адресата в обязательном порядке выделять жирным или подчеркиванием.

Б) Введением дополнительной переменной  $l$  – узнавания ключевых слов из множества  $L$ , которое заранее предусмотрено в предикате зоны описанного выше  $z_i(T, L)$ .

Зоны «Угловой штамп» и «Содержание» однозначно определяются по переменным текста. Для них нет необходимости вводить переменные  $L$ .

Таблица 3 - Ключевые слова (пример)

Значения переменных и их номера  Зона документа $Z=\{z_1, z_2, \dots, z_n\}$	экз.№	Согласовано	Утверждаю	Название до- кумента <sup>1</sup>	Копия:	Директор	Начальник	Исп.	Тел.
	1	2	3	4-19	20	21	22	23	24
Согласование		+							
Утверждение			+						
Заголовок документа				+					
Адресат					+	+	+		
Подпись документа						+	+		
Отметка об исполнителе	+							+	+

Правило построения предиката узнавания зоны документа по ключевым словам текста, выразится следующей формулой:

$$z_i(L) = \bigvee_{\forall b \sim z_i} l_i^b,$$

где  $l_i^b$  – предикат узнавания значения  $b$  ключевого слова соответствующего  $i$ -той зоне.

$$z_{\text{Сог}}^L = l^2; z_{\text{Ут}}^L = l^3; z_{\text{Зд}}^L = l^{4-19}; z_{\text{Адп}}^L = l^{20} \vee l^{21} \vee l^{22}; z_{\text{Под}}^L = l^{21} \vee l^{22}; z_{\text{Исп}}^L = l^1 \vee l^{23} \vee l^{24};$$

Из примера видно, что после ввода дополнительной переменной однозначно определились зоны «Согласование», «Утверждение», «Заголовок документа», «Отметка об исполнителе» имеющий свой уникальный идентификатор. Пара зон «Адресат» и «Подпись документа» несмотря на наличие двух общих переменных соответственно, также однозначно определены т.к. по переменным характеристик текста общих конститuent у них не было.

Правило построения предиката узнавания зоны документа, выразится следующей формулой:

$$z_i(T, L) = \bigwedge \left( \bigwedge_{h=1}^m \bigvee_{a=1}^n t_h^a, \bigvee_{\forall b \sim z_i} l_i^b \right), \quad (1)$$

<sup>1</sup>4-Приказ, 5-Постановление, 6-Регламент, 7-Положение, 8-Инструкция, 9-Донесение, 10-Заявление, 11-Договор, 12-Контракт, 13-План, 14-Отчет, 15-Перечень, 16-Протокол, 17-Акт, 18-Справка, 19-Заявка.

где  $t_h^a$  – предикат узнавания значения  $ah$ -той переменной текста;  $m$  – количество переменных текста,  $n$  – величина алфавита  $h$ -той переменной текста;  $l_i^b$  – предикат узнавания значения  $b$ ключевого слова соответствующего  $i$ -той зоне.

Следовательно, система предикатов позволяющая определить набор имеющихся зон документа будет выглядеть следующим образом:

$$\left\{ \begin{aligned} z_{\text{Уш}}(T, L) &= t_1^1 t_2^1 (t_3^1 \vee t_3^2) t_4^1, \\ z_{\text{Сог}}(T, L) &= (t_1^1 \vee t_1^3) (t_2^1 \vee t_2^2 \vee t_2^3) t_3^1 t_4^2 l^2 = \bar{t}_1^2 \bar{t}_2^3 t_3^1 t_4^2 l^2, \\ z_{\text{Ут}}(T, L) &= t_1^1 t_2^3 t_3^2 t_4^2 l^2, \\ z_{\text{Зд}}(T, L) &= t_1^1 t_2^2 (t_3^1 \vee t_3^2) (t_4^2 \vee t_4^3) l^{4-19} = t_1^1 t_2^2 (t_3^1 \vee t_3^2) \bar{t}_4^1 l^{4-19}, \\ z_{\text{Адп}}(T, L) &= t_1^1 (t_2^2 \vee t_2^3) (t_3^1 \vee t_3^2 \vee t_3^3) t_4^2 (l^{20} \vee l^{21} \vee l^{22}) = t_1^1 (t_2^2 \vee t_2^3) \bar{t}_3^3 t_4^2 (l^{20} \vee l^{21} \vee l^{22}), \\ z_{\text{Сод}}(T, L) &= t_1^2 (t_2^1 \vee t_2^2) (t_3^1 \vee t_3^2 \vee t_3^3 \vee t_3^4) t_4^2 = t_1^2 (t_2^1 \vee t_2^2) t_4^2, \\ z_{\text{Под}}(T, L) &= t_1^3 (t_2^1 \vee t_2^2) (t_3^1 \vee t_3^2) t_4^2 (l^{21} \vee l^{22}), \\ z_{\text{Исп}}(T, L) &= t_1^3 t_2^1 t_3^2 t_4^2 (l^{1} \vee l^{23} \vee l^{24}). \end{aligned} \right.$$

Для определения видов документов проведем анализ их структур по содержащимся в них зонам Таблица 4.

Таблица 4 - Структура построения зон документов (пример)

Зона документа \ Вид документа	Угловой штамп	Подпись документа	Заголовки документа	Адресат	Содержание	Согласование	Утверждение	Отметка об исполнителе
Приказ		+	+		+			+
Постановление	+	+	+		+			+
Регламент		+	+		+	+	+	+
Положение		+	+		+		+	+
Инструкция		+	+		+	+	+	+
Донесение	+	+	+	+	+			+
Заявление		+	+	+	+			
Договор		+	+		+			
Контракт		+	+		+			
План		+	+		+	+	+	+
Отчет	+	+	+	+	+			+
Перечень		+	+		+	+	+	+
Протокол		+	+		+			
Акт		+	+		+		+	+
Справка	+	+	+		+			

Зона документа	Угловой штамп	Подпись документа	Заголовок документа	Адресат	Содержание	Согласование	Утверждение	Отметка об исполнителе
Вид документа								
Служебное письмо	+	+		+	+			+
Заявка	+	+	+	+	+			+

Для удобства восприятия структур документа проведем классификацию документов по группам структур и исключим зону документов «содержание», которое присутствует во всех видах документа и определяющей роли не несет Таблица 5.

Для документов с повторяющимися структурами необходимо ввести дополнительные уникальные переменные (например: наименование документа и т.п.).

В общем случае правило построения предиката узнавания вида документа опишем формулой:

$$v_j(Z, L) = \bigwedge_{\forall z_i=1} \left( \bigwedge_{z_i} z_i, \overline{\bigvee_{z_i=0} z_i, l_j^c} \right), \tag{2}$$

где  $i = \overline{1, n}$ ;  $z_i$  – предикат узнавания требуемой зоны для  $j$ -того вида документа;  $l_j^c$  – предикат узнавания уникального значения ключевого слова  $c$ -того вида документа.

Таблица 5 - Структура используемых зон документов

Зона доку- мента  Вид доку- мента	Угловой штамп	Подпись до- кумента	Заголовок документа	Адресат	Согласова- ние	Утвержде- ние	Отметка об исполнителе	Предикат, описывающий струк- туру документа $v_j(Z, L)$
	$Z_{УШ}$	$Z_{Под}$	$Z_{ЗД}$	$Z_{Адр}$	$Z_{Сог}$	$Z_{Ут}$	$Z_{Исп}$	
С однозначно определенной структурой								
Приказ		+	+				+	$Z_{Под}Z_{ЗД}Z_{Исп}(Z_{УШ} \vee Z_{Адр} \vee Z_{Сог} \vee Z_{Ут})$
Заявление		+	+	+				$Z_{Под}Z_{ЗД}Z_{Адр}(Z_{УШ} \vee Z_{Сог} \vee Z_{Ут} \vee Z_{Исп})$
Справка	+	+	+					$Z_{УШ}Z_{Под}Z_{ЗД}(Z_{Адр} \vee Z_{Сог} \vee Z_{Ут} \vee Z_{Исп})$
Постановле- ние	+	+	+				+	$Z_{УШ}Z_{Под}Z_{ЗД}Z_{Исп}(Z_{Адр} \vee Z_{Сог} \vee Z_{Ут})$
Служебное письмо	+	+		+			+	$Z_{УШ}Z_{Под}Z_{Адр}Z_{Исп}(Z_{ЗД} \vee Z_{Сог} \vee Z_{Ут})$
С повторяющимися структурами								
Договор		+	+					$Z_{Под}Z_{ЗД}(Z_{УШ} \vee Z_{Адр} \vee Z_{Сог} \vee Z_{Ут} \vee Z_{Исп})$
Контракт		+	+					$Z_{Под}Z_{ЗД}(Z_{УШ} \vee Z_{Адр} \vee Z_{Сог} \vee Z_{Ут} \vee Z_{Исп})$
Протокол		+	+					$Z_{Под}Z_{ЗД}(Z_{УШ} \vee Z_{Адр} \vee Z_{Сог} \vee Z_{Ут} \vee Z_{Исп})$
Положение		+	+			+	+	$Z_{Под}Z_{ЗД}Z_{Ут}Z_{Исп}(Z_{УШ} \vee Z_{Адр} \vee Z_{Сог})l^7$
Акт		+	+			+	+	$Z_{Под}Z_{ЗД}Z_{Ут}Z_{Исп}(Z_{УШ} \vee Z_{Адр} \vee Z_{Сог})l^{17}$
Регламент		+	+		+	+	+	$Z_{Под}Z_{ЗД}Z_{Сог}Z_{Ут}Z_{Исп}(Z_{УШ} \vee Z_{Адр})l^6$
Инструкция		+	+		+	+	+	$Z_{Под}Z_{ЗД}Z_{Сог}Z_{Ут}Z_{Исп}(Z_{УШ} \vee Z_{Адр})l^8$
План		+	+		+	+	+	$Z_{Под}Z_{ЗД}Z_{Сог}Z_{Ут}Z_{Исп}(Z_{УШ} \vee Z_{Адр})l^{13}$
Перечень		+	+		+	+	+	$Z_{Под}Z_{ЗД}Z_{Сог}Z_{Ут}Z_{Исп}(Z_{УШ} \vee Z_{Адр})l^{15}$
Донесение	+	+	+	+			+	$Z_{УШ}Z_{Под}Z_{ЗД}Z_{Адр}Z_{Исп}(Z_{Сог} \vee Z_{Ут})l^9$
Отчет	+	+	+	+			+	$Z_{УШ}Z_{Под}Z_{ЗД}Z_{Адр}Z_{Исп}(Z_{Сог} \vee Z_{Ут})l^{14}$
Заявка	+	+	+	+			+	$Z_{УШ}Z_{Под}Z_{ЗД}Z_{Адр}Z_{Исп}(Z_{Сог} \vee Z_{Ут})l^{19}$

Вид документа однозначно определяет его структуру и поиск метаданных необходимо производить не по всему телу документа, а только по зоне, которой в соответствии определен набор метаданных, что значительно сократит время их поиска.

Используя данный подход видно, что возможно описание любого вида документа и создание для любой автоматизированной системы своей уникальной базы знаний для автоматического определения вида документа. Выражения (1) и (2) можно называть правилами формирования формализованных баз знаний автоматизированной системы.

Итак, электронный документ, разработанный согласно ГОСТ Р 6.30 – 2003 и ГОСТ Р ИСО 15489-1— 2007, идентифицируется по множеству реквизитов без введения дополнительных метаданных и методов кластеризации документа. Предложенная электронная модель документа позволяет провести кластеризацию с учетом разработанной системы обеспечения безопасности информации в автоматизированных системах.

#### **Список литературы**

1. М.Ф. Бондаренко, Ю.П. Шабанов-Кушнаренко. Нормальные формы формул алгебры конечных предикатов [Текст]//Научно-технический журнал «Бионика интеллекта». ХНУРЭ, г. Харьков, Украина – 2011 № 3(77).
2. М.Ф. Бондаренко, Ю.П. Шабанов-Кушнаренко. Об алгебре конечных предикатов. [Текст]// Научно-технический журнал «Бионика интеллекта». ХНУРЭ, г. Харьков, Украина – 2011 № 3(77).
3. В.В. Девятков. Системы искусственного интеллекта: Учеб. Пособие для вузов. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2001.
4. Люгер, Джордж, Ф. Искусственный интеллект: стратегия и методы решения сложных проблем, 4-е издание. :Пер. с англ. – М.: Издательский дом «Вильямс», 2003.

#### **References**

1. M.F. Bondarenko, Ju.P. Shabanov-Kushnarenko. Normal'nye formy formul algebry konechnyh predikatov [Tekst]//Nauchno-tehnicheskij zhurnal «Bionika intellekta». HNURJe, g. Har'kov, Ukraina – 2011 № 3(77).
2. M.F. Bondarenko, Ju.P. Shabanov-Kushnarenko. Ob algebre konechnyh predikatov. [Tekst]// Nauchno-tehnicheskij zhurnal «Bionika intellekta». HNURJe, g. Har'kov, Ukraina – 2011 № 3(77).
3. V.V. Devjatkov. Sistemy iskusstvennogo intellekta: Ucheb. Posobie dlja vuzov. – M.: Izd-vo MGTU im. N.Je. Baumana, 2001.
4. Ljuger, Dzhordzh, F. Iskusstvennyj intellekt: strategija i metody reshenija slozhnyh problem, 4-e izdanie. :Per. s angl. – M.: Izdatel'skij dom «Vil'jams», 2003.