

УДК 81'322

## ЗАДАЧИ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТА НА ЕСТЕСТВЕННЫХ ЯЗЫКАХ И ВОЗМОЖНЫЕ МАТЕМАТИЧЕСКИЕ ПОДХОДЫ К ИХ РЕШЕНИЯМ

**А. В. Пруцков**, к.т.н., доцент, доцент кафедры ВПИМ РГРТУ; mail@prutzkow.com

*Рассматривается вопрос математического представления решения задач автоматической обработки текста. Целью работы является анализ используемых математических методов при решении задач автоматической обработки текста и выявление среди них наиболее перспективных для применения при интеллектуальной обработке текста. Существующий математический аппарат, используемый для решения задач автоматической обработки текста, не упрощает это решение и применяется в основном для формализации исследуемой предметной области и процессов в ней. Поэтому актуальным является разработка новых или выявление наиболее перспективных видов математического аппарата для решения указанных задач. Одним из таких видов могли бы стать алгебраические системы. Показано, что все уровни автоматической обработки текста могут быть представлены с помощью алгебраических систем. Приведены примеры из научных работ, содержащих решение задач автоматической обработки текста с помощью различных типов алгебр. Сделан вывод о необходимости сведения операций на уровнях автоматической обработки текста к арифметическим вычислениям.*

**Ключевые слова:** автоматическая обработка текста, математика, алгебра, морфологический уровень, синтаксический уровень, семантический уровень.

### Введение

Несмотря на динамичное развитие автоматической обработки текста, при решении многих её задач остро ощущается отсутствие типовых методов решения. Это приводит к необходимости разрабатывать подходящий аппарат для описания каждого метода решения. Математическое обеспечение, доступное исследователю в данный момент, не обеспечивает всех потребностей, а значит, не упрощает решение задач автоматической обработки текста в достаточной степени. Поэтому необходимо проанализировать задачи автоматической обработки текста, существующий математический аппарат и выявить наиболее перспективный инструментарий для решения задач автоматической обработки текста, что и будет сделано в данной статье.

### Автоматическая обработка текстов и её задачи

В широком смысле автоматическая обработка текста представляет собой преобразование текста с помощью ЭВМ [1].

Все задачи автоматической обработки текста можно разбить на две группы:

- 1) задачи поверхностной обработки текста;
- 2) задачи интеллектуальной обработки текста.

В задачах поверхностной обработки текст рассматривается исключительно как совокупность символов (или их числовых кодов в определенном стандартном способе кодирования). Примерами задач этой группы являются хранение, редактирование, форматирование (придание элементам текста определенных стилей) и другая полиграфическая обработка (кроме автоматизированной корректуры).

В задачах интеллектуальной обработки текст рассматривается как носитель человеческих знаний. «По имеющимся в литературе оценкам, более 70 % ресурсов [знаний], накопленных в различных организациях, носит неструктурированный характер и образуется электронными текстовыми документами» [2]. Примерами задач интеллектуальной обработки текста являются задачи проверки правописания и грамматики, понимания смысла текста. Все задачи этой группы являются интеллектуальными. Интеллектуальными будем называть задачи, которые не имеют общепринятых, «классических» методов решения. Для решения интеллектуальной задачи необходимо разработать разрешающий ее алгоритм.

Задачи первой группы имеют решения, реализованные в виде алгоритмов в программных системах. Задачи второй группы представляют научный интерес, так как не имеют общеприня-

тых методов решения. Поэтому далее под термином «автоматическая обработка текста» в узком смысле будем понимать интеллектуальную автоматическую обработку текста.

По мнению автора, наиболее перспективными задачами второй группы следует считать следующие:

- анализ текста и выявление в нём значимых [3];
- диалог ЭВМ с пользователем на естественном языке [4];
- классификация текстов по категориям на основе статистики, полученной в ходе морфологического анализа текстов [5];
- использование нечетких множеств, нечеткой логики и нейронных сетей для интеллектуальной обработки текста [6, 7].

При решении большинства перечисленных, а также других задач интеллектуальной обработки текста научные результаты, применимые на практике, получены лишь для ограниченного естественного языка.

#### **Уровни интеллектуальной автоматической обработки текстов**

При интеллектуальной автоматической обработке текста решаются две основные задачи:

- 1) анализ текста с целью выявления его морфологических, синтаксических и семантических параметров;
- 2) синтез текста с необходимыми семантическими, синтаксическими и морфологическими параметрами.

Перечисленные виды параметров позволяют выделить три уровня автоматической обработки текста, каждый из которых решает соответствующие задачи.

Функцией морфологического уровня является соотношение словоформ и грамматических значений.

Функцией синтаксического уровня является соотношение слов и их ролей в предложении, связи слов в предложении и тексте.

Функцией семантического уровня является соотношение словосочетаний и предложений и их смыслового значения.

#### **Математическое обеспечение интеллектуальной автоматической обработки текста**

Несмотря на важность и актуальность (вот уже более 50 лет) автоматической обработки текста, а также появления такой науки, как математическая лингвистика, для этой научной сферы не предложено специализированного математического аппарата, сделавшего подходы к решению задач автоматической обработки текста общепринятыми.

Например, для описания структуры предложения и текста при синтаксическом анализе и синтезе могут быть использованы формальные грамматики, расширенные сети переходов, структурные формулы в том или ином представлении и другие подходы. Однако наличие такого большого числа подходов свидетельствует о том, что они имеют определенные преимущества и недостатки, которые ограничивают их область применения и не делают их универсальными.

Для решения задач автоматической обработки текста можно было бы использовать формальные языки. Однако они неадекватны естественным языкам и широко для решения задач автоматической обработки текста не используются.

Одной из причин такой ситуации в этой научной области является то, что законы естественного языка порождены человеком и могут изменяться как под воздействием внешних воздействий (например, под влиянием культуры одного народа на культуру другого), так и законодательно. Естественный язык изначально был неформализованным, и почти все его изменения также имели неформализованный вид. Лишь филологи пытались формализовать язык в виде грамматических правил, имевших при этом большое число исключений.

Поэтому построение любых математических моделей для решения задач автоматической обработки текста представляет трудность, так как язык является неформализованным объектом.

В этих условиях исследователям в области автоматической обработки текста необходимо:

- использовать как можно более абстрактный математический аппарат;
- «подгонять» существующий математический аппарат для решения их научных задач.

Это вызывает критику в их адрес как со стороны математиков, так и со стороны филологов. Таким образом, автоматическая обработка текста с точки зрения применяемого математического аппарата превращается в «поскутное одеяло».

#### **Алгебраические системы**

Одним из перспективных видов математического аппарата, используемого при решении задач автоматической обработки текста, являются алгебраические системы.

Алгеброй  $A$  называют множество некоторых объектов  $M$  (носитель) и множество операций над ними  $\Omega$  (сигнатура) [8]:

$$A = (M, \Omega).$$

Математический аппарат алгебры универсален и применяется для решения самых различных задач. Также алгебры можно применить для ре-

шения задач автоматической обработки текста.

Рассмотрим уровни автоматической обработки текста и виды алгебр, применяемых для решения задач этих уровней.

### Морфологический уровень

В работе [9] предложена модель, описывающая формобразование для генерации и определения форм слов различных естественных языков.

Для описания модели предложена алгебраическая система  $E = \{\Psi, \Omega\}$ , где  $\Psi$  – множество слов естественного языка,  $\Omega$  – множество операций над словами [10].

Множество  $\Psi = S_M \cup F_M \cup \{\emptyset\}$ , где  $S_M$  – множество нормальных форм слов;  $F_M$  – множество остальных форм слов;  $S_M \cap F_M = \emptyset$ ;  $\emptyset$  – пустое слово.

Пусть  $x \in X$ ,  $y \in Y$ ,  $Z = X \times Y$ . Через  $Im(x, Y)$  обозначен образ  $x$  в  $Y$  при соответствии  $Z$ , а через  $pIm(y, X)$  – прообраз  $y$  в  $X$  при соответствии  $Z$ .

Соответствие  $L \subset S_M \times F_M$  обладает следующими свойствами:

$$1) |Im(S, F_M)| \geq 0, \text{ где } S \in S_M;$$

$$2) |pIm(F, S_M)| = 1, \text{ где } F \in F_M;$$

$$3) \forall S_i \forall S_j (Im(S_i, F_M) \cap Im(S_j, F_M) = \emptyset), \text{ где } S_i, S_j \in S_M; i, j = 1, 2, \dots, N_S; i \neq j; N_S = |S_M|.$$

Множество  $\Omega = \Theta_M \cup \Theta'_M$ , где  $\Theta_M$  – множество прямых операций;  $\Theta'_M$  – множество обратных операций;  $\Theta_M \cap \Theta'_M = \emptyset$ ;  $|\Theta_M| = |\Theta'_M| = N_\Theta$ .

Множество  $T = \{(\Theta_i, \Theta'_i) \mid i = 1, 2, \dots, N_\Theta\} \subset \Theta_M \times \Theta'_M$  – взаимно однозначное соответствие.

Операции множества  $\Omega$  в дальнейшем будем называть преобразованиями.

Пусть  $Q \in \Omega$  – преобразование, а  $Q' \in \Omega$  – преобразование, обратное данному. Пара преобразований  $Q$  и  $Q'$  обладает следующим свойством:

$$[(Q, Q') \in T] \otimes [(Q', Q) \in T] = 1,$$

где  $\otimes$  – логическая операция сложения по модулю 2 (исключающее «ИЛИ»); 1 – логическая константа «ИСТИНА».

Цепочка преобразований (прямая цепочка преобразований) – это конечная упорядоченная последовательность (вектор) преобразований:

$$R = (Q_1, Q_2, \dots, Q_n) \in R_M,$$

где  $Q_1, Q_2, \dots, Q_n \in \Omega$ ;  $n \geq 1$ ;  $R_M$  – множество цепочек.

Обратная цепочка преобразований  $R'$  представляет собой обратную последовательность преобразований, обратных данным:

$$R' = (Q'_n, Q'_{n-1}, \dots, Q'_1) \in R_M,$$

где  $Q'_1, Q'_2, \dots, Q'_n \in \Omega$ .

Каждая цепочка преобразований должна обладать следующими свойствами:

1) однозначность результата: цепочка всегда приводит к одному и тому же результату;

2) обратимость действия: применение к форме  $A$  прямой цепочки, а затем обратной цепочки не изменяет ее:

$$A = (A(R))(R'),$$

где  $A(R)$  – результат применения к форме  $A$  цепочки преобразований  $R$ .

Данное представление формобразования позволило разработать метод, позволяющий генерировать и определять формы слов естественных языков различных семейств и групп [11]. Метод реализован в системе автоматизированного обучения морфологии естественных языков [12].

### Синтаксический уровень

В работе [8] приведены структурные формулы словосочетаний. Некоторые из словосочетаний приведены в таблице. Нумерация формул сохранена. В таблице используются следующие обозначения. Заглавные буквы: С – существительные; П – прилагательные, Р – предлоги. Строчные буквы используются для обозначения падежа, например р – родительный.

Таблица

№ п/п	Структурная формула	Словосочетание-представитель
1	ПС	Индикаторное устройство
...	...	...
3	ПППС	Управляющая цифровая вычислительная машина
4	С → Ср	Испытания машин
...	...	...
9	ПС → Ср → Ср	Автоматизированная система поиска информации
...	...	...
15	ПС → Р → Ср → Ср	Символические языки для поиска информации
...	...	...

Для описания свойств объектов, отношений и зависимостей между ними можно использовать алгебру предикатов первого порядка (это будет также продемонстрировано на примере семантического уровня). Введем предикат, определяющий зависимость слова Б от слова А в перечисленных выше структурных формулах:

*Зависимость (А, Б, падеж, положение).*

Здесь параметр *положение* определяет порядок следования А и Б и принимает два значения: *до* – если Б предшествует А и *после* – если Б следует за А. Параметр *падеж* определяет падеж Б и принимает значение из множества {и, р, д, в,

$t, n, l\}$ , где  $u, p, d, v, t, n$  – первые буквы названия падежей русского языка;  $l$  – значение, обозначающее любой падеж, то есть параметры не зависят от падежа.

Структурные формулы из таблицы можно описать с помощью предиката *Зависимость* следующим образом:

$$P_1 = \text{Зависимость}(C, П, л, до);$$

...

$$P_3 = \text{Зависимость}(C, П_1, л, до) \wedge$$

$$\wedge \text{Зависимость}(C, П_2, л, до) \wedge$$

$$\wedge \text{Зависимость}(C, П_3, л, до);$$

$$P_4 = \text{Зависимость}(C_1, C_2, p, после);$$

...

$$P_9 = \text{Зависимость}(C_1, П, л, до) \wedge$$

$$\wedge \text{Зависимость}(C_1, C_2, p, после) \wedge$$

$$\wedge \text{Зависимость}(C_2, C_3, p, после);$$

...

$$P_{15} = \text{Зависимость}(C_1, П, л, до) \wedge$$

$$\wedge \text{Зависимость}(C_1, P, л, после) \wedge$$

$$\wedge \text{Зависимость}(C_1, C_2, p, после) \wedge$$

$$\wedge \text{Зависимость}(C_2, C_3, p, после).$$

Здесь  $\wedge$  – логическая операция конъюнкция.

Таким образом, структурные формулы описания словосочетаний записаны с помощью алгебры предикатов.

При синтаксическом анализе слова предложения в виде предикатных формул и предикаты, описывающие структурные формулы, могут сопоставляться с помощью теории унификации [14].

Алгебра предикатов для решения задач синтаксического уровня автоматической обработки текста также используется в работе [15].

### Семантический уровень

В работе [2] используется как алгебра предикатов вместе с логическими операциями и кванторами, так и новые классы формальных объектов, называемые концептуальными базами.

Концептуальный базис – это упорядоченный набор вида

$$((c_1, \dots, c_4), (c_5, \dots, c_8), (c_9, \dots, c_{15})),$$

где  $c_1, \dots, c_{15}$  – конечные или счетные множества символов и выделенные элементы таких множеств.

Некоторые из составляющих концептуального базиса:  $c_1 = St$  – конечное множество символов, называемых сортами и обозначающих наиболее общие рассматриваемые понятия;  $c_3 = Gen$  – бинарное отношение на  $St$ .

Автор отмечает, что отдаленным прообразом концептуального базиса является сигнатура алгебраической системы.

Каждому концептуальному базису  $B$  ставятся в соответствие три класса формул  $Ls = Ls(B)$ ,  $Ts = Ts(B)$ ,  $Ys = Ys(B)$  ( $l$ -формулы,  $t$ -формулы,  $y$ -

формулы соответственно), используемые для построения семантических представлений с помощью правил построения  $P[0], \dots, P[10]$ .

Приведем пример семантического представления предложения «В каждой стране Европы есть город с количеством жителей, превышающим 30 тысяч человек» с использованием предикатов, кванторов и логических операций:

$$\forall x_1 (\text{страна}^*(\text{Место}, \text{Европа}) \exists x_2 (\text{город}(\text{Место}(x_2, x_1) \wedge \text{Меньше}(30000, \text{Колич.элемент}(\text{Жители}(x_2))))),$$

где  $\forall$  – квантор общности,  $\exists$  – квантор существования.

Пример семантического представления составных объектов в предложении «П. Сомов знает, что И. Семенов является директором фирмы, персонал которой включает 38 человек»:

$$\text{Знает}(\text{П. Сомов}, \text{Сейчас}, ((\text{И. Семенов} \equiv \text{Директор}(\text{нек фирма}^*(\text{Описание}, P1): x1)) \wedge (P1 \equiv (\text{Колич}(\text{Персонал}(x1)) \equiv 38))))).$$

В работе [2] синтаксический и семантический уровни объединены в один уровень.

Алгебра предикатов используется и в работе [16]. Здесь предикаты используются для сопоставления существительных и элементов семантического значения  $y$ , признака одушевленности  $x$ , грамматического падежа существительных  $z$ . Примерами элементов семантических значений в работе [16] являются  $y^m$  – механизм;  $y^c$  – имя собственное;  $y^i$  – инструмент;  $y^t$  – часть тела;  $y^p$  – плоскость/точка;  $y^o$  – объемное пространство;  $y^b$  – определенное время;  $y^p$  – период;  $y^n$  – пункт назначения.

Семантический падеж существительного предложения задается предикатом  $P(x, y, z)$ , который определяется авторами как:

$$P(x, y, z) = P(x) \wedge P(y) \wedge P(z).$$

Необходимость разработки адекватного математического аппарата особенно актуальна на семантическом уровне. Отсутствие такого аппарата вынуждает исследователей в этой сфере автоматической обработки текста работать с ограниченным естественным языком.

Несмотря на наличие таких моделей представления знаний, как семантические сети, фреймы, продукционные системы и логические модели, не существует универсальных алгоритмов, преобразующих смысл в текстовом виде к одной из перечисленных моделей. Применение таких алгоритмов позволило бы отойти от текста как способа хранения знаний к формализованным семантическим представлениям.

### Заключение

Из всего вышеизложенного можно сделать

следующие выводы.

1. Поставлена проблема необходимости разработки математического аппарата для решения задач интеллектуальной автоматической обработки текста.

2. Показано, что задачи автоматической обработки текста могут быть решены с помощью теории алгебр. Алгебраические системы для решения задач автоматической обработки текста на различных уровнях имеют различные носители и сигнатуры, что усложняет решение задач.

3. На синтаксическом и семантическом уровнях может использоваться алгебра предикатов для наделения объектов определенными свойствами. Наличие у предикатов области значений, включающей всего два элемента («ИСТИНА» и «ЛОЖЬ»), позволяет описывать с помощью предикатов наличие или отсутствие свойств у объектов исследуемой предметной области.

4. На морфологическом уровне необходимо выполнять операции над словами для получения основ или словоформ. Это требует использования алгебры со свойствами, отличными от свойств алгебры предикатов.

5. Алгебраические системы представляют собой абстрактное описание. Однако при практическом использовании их сигнатуры должны включать уже конкретные операции. Поэтому целесообразно свести интеллектуальную обработку текстов к арифметическим вычислениям. Например, на морфологическом уровне генерация и определение словоформ могли бы сводиться к двум функциям:

$$\begin{aligned} 1) f &= g(s); \\ 2) s &= g^{-1}(f), \end{aligned}$$

где  $f$  – числовое представление словоформы  $F$  – элемента множества  $F_M$ ;  $s$  – числовое представление основы  $S$  – элемента множества  $S_M$ ;  $y = g(x)$  – функция, вычисляющая по числовому представлению основы числовое представление словоформы;  $y = g^{-1}(x)$  – обратная функция к  $g(x)$ .

Операции на остальных уровнях автоматической обработки текста также необходимо свести к вычислениям. При реализации такого сведения рассматриваемые задачи автоматической обработки текста перестанут быть интеллектуальными, так как получат математическое решение.

#### Библиографический список

1. **Языкознание.** Бол. энцикл. словарь / гл. ред. В.Н. Ярцева. 2-е изд. М.: Бол. рос. энцикл., 1998. 685 с.
2. **Фомичев В. А.** Формализация проектирования лингвистических процессоров. М.: Макс Пресс, 2005. 368 с.
3. **Люгер Дж. Ф.** Искусственный интеллект:

стратегии и методы решения сложных проблем. 4-е изд.: пер. с англ. М.: Издательский дом «Вильямс», 2003. 864 с.

4. **Правиков А. А., Фомичев В. А.** Разработка рекомендательной системы с естественно-языковым интерфейсом на основе математических моделей семантических объектов // Бизнес-информатика. Междисциплинарный науч.-практ. журнал ГУ-ВШЭ. 2010. № 4 (14). С. 3-11.

5. **Токарев В. Л., Трутнев Е. В.** Автоматизированная классификация текстов Интернет-страниц по функциональным стилям речи на основе морфологического анализа // Известия Тульского государственного университета. Технические науки. 2008. Вып. 4. С. 249-256.

6. **Громов Ю. Ю., Поляков Д. В., Авдеева Т. О.** Формализация текстовой коллекции на основе нечетких частот коллокаций // Приборы и системы. Управление, контроль, диагностика. 2013. № 2. С. 15-17.

7. **Третьяков Ф. И., Серебряная Л. В.** Модель универсальной десятичной классификации текстов с использованием искусственной нейронной сети // Проблемы передачи и обработки информации в сетях и системах телекоммуникаций: материалы 18-й Международ. науч.-техн. конф. М.: Горячая линия – Телеком, 2015. С. 208-211.

8. **Кузнецов О. П., Адельсон-Вельский Г. М.** Дискретная математика для инженера. 2-е изд., перераб и доп. М.: Энергоатомиздат, 1988. 480 с.

9. **Пруцков А. В.** Морфологический анализ и синтез текстов посредством преобразований форм слов // Вестник Рязанской государственной радиотехнической академии. 2004. № 15. С. 70-75.

10. **Пруцков А. В.** Алгебраическое представление модели формобразования естественных языков // Cloud of Science. 2014. Т. 1. № 1. С. 119-128.

11. **Prutskov A. V.** Algorithmic Provision of a Universal Method for Word-Form Generation and Recognition // Automatic Documentation and Mathematical Linguistics, 2011, Vol. 45, No. 5, pp. 232-238.

12. **Пруцков А. В.** Статический и динамический подходы к проектированию подсистем проверки знаний автоматизированных обучающих систем // Информационные ресурсы России. 2006. № 1. С. 27-29.

13. **Белоногов Г. Г., Богатырев В. И.** Автоматизированные информационные системы / под ред. К. В. Тараканова. М.: Сов. радио, 1973. 328 с.

14. **Каширин Д. И., Каширин И. Ю., Пылькин А. Н.** Полиморфическое представление знаний в Semantic Web: монография. М.: Горячая линия – Телеком, 2009. 136 с.

15. **Борисова Н. В., Канищева О. В.** Моделирование синтаксического анализа в задачах аннотирования и реферирования полнотекстовых документов // Сборник научных трудов «Вестник НТУ "ХПИ"». 2009. № 4. С. 87-96.

16. **Хайрова Н., Шаронова Н.** Логико-лингвистическая модель извлечения фактов из слабоструктурированной текстовой информации // International Journal «Information Models and Analyses». 2013. Vol. 2. No 3. pp. 167-175.

UDC 81'322

## NATURAL LANGUAGE PROCESSING PROBLEMS AND POSSIBLE APPROACHES TO THEIR SOLUTION

**A. V. Prutzkow**, PhD (technical sciences), associate professor, RSREU, Ryazan; mail@prutzkow.com

*We consider the subject of mathematical representation of natural language processing problems. The aim of the article is to analyse current mathematical methods for natural language processing problem solution and expose what methods are prospective in artificial intelligence language processing. Current mathematical methods do not make easy the solution of natural language processing problems and are used for formalization of problem area research and processes in it. So to develop new problem solution methods and to reveal prospective ones are actual tasks nowadays. One of such methods is an algebraic system. We demonstrate all natural language processing levels represented by algebraic system. We give different scientific work examples of natural language processing problem solution with the help of different algebra types. We conclude that natural language processing operation is necessary to be transformed into arithmetic calculations.*

**Keywords:** natural language processing, mathematics, algebra, morphological level, syntactical level, semantic level.

### References

1. **Jazykoznanie.** *Bol. ehncikl. slovar'* (Linguistics. Big encyclopaedic dictionary). gl. red. V.N. Jarceva. 2-e izd. Moscow: Bol. ros. ehncikl., 1998, 685 p. (in Russian).
2. **Fomichev V. A.** *Formalizacija proektirovanija lingvisticheskikh processorov* (Linguistic processor designing foramlisation). Moscow: Maks Press, 2005, 368 p. (in Russian).
3. **Ljoger Dzh. F.** *Iskusstvennyj intellekt: strategii i metody reshenija slozhnykh problem* (Artificial intelligence: strategies and comlex problem solutions). 4th ed.: per. s angl. Moscow: Izdatel'skij dom «Vil'jam», 2003, 864 p. (in Russian).
4. **Pravikov A. A., Fomichev V. A.** Razrabotka rekomendatel'noj sistemy s estestvenno-jazykovym interfejsom na osnove matematicheskikh modelej semanticheskikh ob"ektov. *Biznes-informatika. Mezhdisciplinarnyj nauch.-prakt. zhurnal GU-VSHEH*. 2010, no. 4 (14), pp. 3-11 (in Russian).
5. **Tokarev V. L., Trutnev E. V.** Avtomatizirovannaja klassifikacija tekstov Internet-stranic po funkcional'nym stiljam rechi na osnove morfologicheskogo analiza. *Izvestija Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki*. 2008, no. 4, pp. 249-256 (in Russian).
6. **Gromov Ju. Ju., Poljakov D. V., Avdeeva T. O.** Formalizacija tekstovoj kollekcii na osnove nechetkikh chastot kollokacij. *Pribory i sistemy. Upravlenie, kontrol', diagnostika*. 2013, no. 2, pp. 15-17 (in Russian).
7. **Tretyakov F. I., Serebryanya L. V.** Model universalnoj desjaticnoy klassifikatsii tekstov s ispol'zovaniem iskusstvennoy neyronnoj seti. Problemy peredachi i obrabotki informatsii v setjah i sistemah telekommunikatsii: materialy 18-y Mezhdunar. nauch.-tehn. konf. Moscow: Gorjachaja linija – Telekom, 2015. pp. 208-211.
8. **Kuznecov O. P., Adel'son-Vel'skij G. M.** *Diskretnaja matematika dlja inzhenera* (Descrete mathematics for engineer). 2nd ed., pererab i dop. Moscow: Ehnergoatomizdat, 1988, 480 p. (in Russian).
9. **Prutzkow A. V.** Morfologicheskij analiz i sintez tekstov posredstvom preobrazovanij form slov. *Vestnik Rjazanskoj gosudarstvennoj radiotekhnicheskoy akademii*. 2004, no. 15, pp. 70-75 (in Russian).
10. **Prutzkow A. V.** The algebraic representation of the natural language wordform building model. *Cloud of Science*. 2014, vol. 1, no. 1, pp. 119-128 (in Russian).
11. **Prutzkow A. V.** Algorithmic Provision of a Universal Method for Word-Form Generation and Recognition. *Automatic Documentation and Mathematical Linguistics*. 2011, vol. 45, no. 5, pp. 232-238.
12. **Prutzkow A. V.** Staticeskij i dinamicheskij podkhody k proektirovaniju podsistem proverki znanij avtomatizirovannykh obuchajushchikh system. *Informacionnye resursy Rossii*. 2006, no. 1, pp. 27-29 (in Russian).
13. **Belonogov G. G., Bogatyrev V. I.** *Avtomatizirovannye informacionnye sistemy* (Automatized inforamtion systems). ed. K. V. Tarakanov. Moscow: Sov. radio, 1973, 328 pp. (in Russian).
14. **Kashirin D. I., Kashirin I. Ju., Pyl'kin A. N.** *Polimorficheskoe predstavlenie znanij v Semantic Web: monografija* (Polymorphic representation of knowledges in Semantic Web). Moscow: Gorjachaja linija – Telekom, 2009, 136 p. (in Russian).
15. **Borisova N. V., Kanishcheva O. V.** Modelirovanie sintaksicheskogo analiza v zadachakh annotirovanija i referirovanija polnotekstovykh dokumentov. *Sbornik nauchnykh trudov «Vestnik NTU "KHPI"»*. 2009, no. 4, pp. 87-96 (in Russian).
16. **Khajrova N., Sharonova N.** Logiko-lingvisticheskaja model' izvlechenija faktov iz slabostrukturirovannoj tekstovoj informacii. *International Journal «Information Models and Analyses»*. 2013, vol. 2, no. 3, pp. 167-175 (in Russian).