

УДК 004.89

doi: 10.26102/2310-6018/2018.23.4.031

С.С. Васильев, А.А. Харитонов, Д.М. Коробкин, С.А. Фоменков.  
**ИЗВЛЕЧЕНИЕ ОПИСАНИЙ МОРФОЛОГИЧЕСКИХ  
ПРИЗНАКОВ ТЕХНИЧЕСКИХ ОБЪЕКТОВ ИЗ РУССКОЯЗЫЧНЫХ  
ПАТЕНТОВ**

*Волгоградский государственный технический университет,  
Волгоград, Россия*

*Задача автоматизации синтеза инновационных решений в области технических систем и технологий является одной из наиболее приоритетных проблем науки. Авторы работы предлагают осуществлять автоматизацию важнейших, начальных этапов проектирования новых технических систем и технологий на основе актуализируемых баз знаний, полученных из мирового патентного массива, в том числе из патентной базы РосПатента. Согласно методу морфологического анализа и синтеза предполагается, что из некоторого технического решения (патента) извлекаются основные структурные признаки (функции технических объектов), на основе которых ищутся альтернативные варианты. Все полученные признаки сводятся в морфологическую таблицу, комбинируются, что дает множество новых решений. В данной работе описывается процесс разработки модуля извлечения описаний технических функций из российских патентов. Была сформирована грамматика представления описаний технических функций в текстах русскоязычных патентов согласно модели «Действие-Объект-Условие»; разработаны алгоритмы первичной обработки патентного массива, извлечения технических функций посредством анализа деревьев зависимостей, формирования морфологической таблицы. Работоспособность программного модуля, состоящего из блока обработки патентного массива; блока сегментации текста патентной формулы; блока семантического анализа текста; блока извлечения описаний технических функций; блока представления результатов обработки патентного массива, была проверена на ряде тестовых задач.*

**Ключевые слова:** технические функции, обработка естественно-языковых текстов, патенты, РосПатент, Link Grammar Parser, грамматика

### **Введение**

Задача автоматизации синтеза инновационных решений [1,2] в области технических систем и технологий является одной из наиболее приоритетных проблем науки. Авторы работы предлагают [3,4] осуществлять автоматизацию важнейших, начальных этапов проектирования новых технических систем и технологий на основе актуализируемых баз знаний, полученных из мирового патентного массива [5,6], в том числе из патентной базы РосПатента [7].

Результатом синтеза изобретения является новое проектное решение с новыми свойствами, которые не встречались ранее. Структурный синтез [1] подразумевает существование некоторой технической функции, для которой необходимо определить некий объект или систему, которые

реализуют данную техническую функцию при определенных условиях и ограничениях. По методу морфологического анализа и синтеза [1] предполагается, что из некоторого технического решения извлекаются основные структурные признаки, на основе которых ищутся альтернативные варианты. Все полученные признаки сводятся в морфологическую таблицу, комбинируются, что дает множество новых решений. Одним из разновидностей морфологического анализа является подход [8,9], при котором в качестве признаков берутся функции отдельных объектов (его элементов). В таком случае, альтернативами являются различные реализации этих функций.

Алгоритм морфологического анализа и синтеза содержит следующие последовательные процедуры:

- а) выделение основных признаков объекта;
- б) подготовка формы для заполнения морфологической таблицы;
- в) заполнение таблицы альтернативами;
- г) выбор наилучших вариантов решений;
- д) формирование технического решения.

**Цель работы** - осуществить извлечение технических функций из патентов ФИПС РосПатента и представить их в виде морфологической таблицы, являющейся основой процедуры синтеза новых проектных решений.

#### **Материалы и методы**

##### ***Алгоритм анализа патентного массива***

По результатам проведенного исследования, заключавшегося в анализе предметной области патентования изобретений, знакомстве со структурой патента, проведении лингвистического анализа патентной формулы, обзора существующих автоматизированных методов синтеза новых технических решений, исследовании этапов автоматической обработки текста, был предложен алгоритм анализа патентного массива.

На вход алгоритма поступает папка с патентной выборкой, содержащая данные о патентах, выданных российским патентным ведомством. Из текстов патентных формул извлекаются технические функции, в которых глаголы определяют основное действие и которые будут однозначно описывать сущность изобретения. Затем определяются классы действий глаголов, формируя тем самым обобщенные технические функции, а сами функции записываются в варианты реализации (альтернативы) выделенных обобщенных функций.

Выходными данными алгоритма является морфологическая таблица, представленная визуально на программной экранной форме и наглядно отображающая варианты исполнения (альтернативы) обобщенной

технической функции. Полученная морфологическая таблица может быть использована для поиска основополагающей идеи для синтеза нового технического решения.

### *Извлечение описаний технических функций из патентов*

Из xml-файла патента РосПатента извлекается следующая информация:

- `<B110>... </B110>` - номер патента;
- `<B220><date>... </date></B220>` - дата выдачи патента;
- `<ru-b542>... </ru-b542>` - название патента;
- `<ru-b560>... </ru-b560>` - ссылки на другие патенты являющиеся источниками вдохновения;
- `<B721><ru-name-text>... </ru-name-text></B721>` список авторов изобретения;
- `<claims>... </claims>` - формула изобретения.
- 

Непосредственно текст патентной формулы заключен в теги `<claim-text></claim-text>`. Формулы патентов строятся по одному шаблону, отличному от обычной структуры предложения. Чаще всего формула представляет собой одно сложноподчиненное предложение, в котором несколько придаточных, каждое из которых дополняет свойства объекта, о котором говорится в главном предложении. Синтаксический разбор такого длинного предложения будет очень затратной по времени и ресурсам процедурой. Чтобы уменьшить число ошибок в результатах семантического разбора, данное предложение делится на несколько частей (сегментируется), а затем каждая часть исходного предложения разбирается отдельно.

Алгоритм сегментации [10] состоит из следующих последовательных процедур:

- а) из предложения удаляются словосочетания «по п.» и «по пункту».
- б) на символ переноса строки заменяются все склонения словосочетаний: «отличающийся тем, что»; «предусматривающий»; «обеспечивающий»; «включающий»; «который»; «чей»; также удаляются все предлоги, относящиеся к данным словам.
- в) заменяются на символ переноса строки словосочетания: «а»; «но»; «где»; «причем»; «при этом»; «путем»; «так, что»; «кроме того»; «перед»; «после».
- г) удаляются все знаки вопросов, двоеточия, и точки с запятой.

д) удаляются символы обозначающие списки, например, «а)» , «1.» и так далее.

е) убираются все многоточия, и излишние повторения символов переноса строки и запятых.

Для выполнения семантического разбора сегмента патентной формулы используется программное обеспечение Link Grammar Parser [11], основанный на грамматике связей. Парсер принимает на вход сегменты предложения на естественном языке, и выдает связи, найденные в предложении, с разметкой морфологических признаков слов сегмента. Основная идея грамматики связей – представление слов как блоков с исходящими от них соединителями. Соединители могут указывать вправо («+») или влево («-»). Два соединителя, один из которых указывает влево и связывается с соединителем, указывающим вправо, образуют связь. На выходе получаем не синтаксические связи (субъект/объект), а отношения между парами слов [12].

Для разбора сегментов патентных формул понадобятся следующие типы связей Link Grammar Parser [11]:

- W – соединяет границу предложения и главное слово, как правило являющееся существительным;
- MV – соединяет глаголы с их объектами;
- SI – соединяет глаголы с их объектами отвечающими на вопрос что;
- E – соединяет наречие с глаголом;
- J – соединяет предлог с зависимым существительным или местоимением;
- I – соединяет глагол с другим зависимым от него глаголом;
- A – соединяет местоимение или прилагательное с существительным;
- M – соединяет существительные с другими зависимыми от них существительными;
- EI – соединяет глагол с наречием;
- PI – соединяет отглагольные прилагательные или причастия с существительными;
- AXP – соединяет местоимения или прилагательные с прилагательными

### Контекстно-зависимая грамматика представления технической функции

Описание функции любого технического объекта будем представлять в виде тройки множеств, представленных формулой [2]:

$$F = \langle D, G, H \rangle \quad (1)$$

где  $D$  – множество действий, производимых и приводящих к желаемому результату;

$G$  – множество объектов (операндов), на которые эти действия направлены;

$H$  – множество особых условий и ограничений выполняемых действий.

Разбор множества предложений показал, что сегменты патентных формул имеют сходную синтаксическую структуру. Это наблюдение позволило составить контекстно-зависимую грамматику (2) для извлечения компонентов модели (1) из предложений патентных формул посредством анализа связей:

$$Gram = (T, N, \langle W \rangle, R), \quad (2)$$

где  $T = \{\text{действие, объект, условие}\}$  - набор терминалов,

$N = \{\langle MV \rangle, \langle SI \rangle, \langle E \rangle, \langle J \rangle, \langle I \rangle, \langle A \rangle, \langle M \rangle\}$  - набор не-терминалов,  
 $\langle W \rangle$  - начальный не-терминал [13],

$R$ - набор продукционных правил:

$\langle W \rangle \rightarrow \text{действие} \langle MV \rangle \langle E \rangle | \text{действие} \langle MV \rangle | \text{действие} \langle E \rangle | \text{действие} \varepsilon$   
 $\text{действие} \langle MV \rangle \rightarrow \text{действие} \langle MV \rangle | \text{действие} \langle MV \rangle \langle MV \rangle$   
 $\text{действие} \rightarrow \text{действие} | \langle I \rangle \text{действие} | \text{действие} \langle I \rangle$   
 $\langle I \rangle \text{действие} \rightarrow \text{действие} \text{действие} | \langle I \rangle \text{действие} \text{действие} | \text{действие} \langle I \rangle$   
 $\text{действие} \langle I \rangle \rightarrow \text{действие} \text{действие} | \text{действие} \text{действие} \langle I \rangle | \text{действие} \langle I \rangle$   
 $\text{действие}$   
 $\langle MV \rangle \rightarrow \text{объект} \langle MV \rangle | \langle A \rangle \text{объект} | \langle A \rangle \text{объект} \langle M \rangle | \text{объект} \langle M \rangle | \text{объект}$   
 $\langle A \rangle \text{объект} \rightarrow \text{объект} \text{объект} | \langle A \rangle \text{объект} \text{объект} | \text{объект} \langle A \rangle \text{объект}$   
 $\text{объект} \langle M \rangle \rightarrow \text{объект} \text{объект} \text{объект} | \text{объект} \langle A \rangle \text{объект} | \text{объект} \text{объект} \langle M \rangle |$   
 $\text{объект} \langle A \rangle \text{объект} \langle M \rangle$   
 $\text{условие} \langle J \rangle \rightarrow \text{условие} | \text{условие} \langle E \rangle | \text{условие} \langle J \rangle$   
 $\langle A \rangle \text{условие} \rightarrow \text{условие} \text{условие} \langle M \rangle | \text{условие} \langle A \rangle \text{условие} \langle M \rangle$   
 $\text{условие} \langle M \rangle \rightarrow \text{условие} \text{условие} | \text{условие} \text{условие} \langle M \rangle | \text{условие} \text{условие} \langle J \rangle |$   
 $\text{условие} \langle A \rangle \text{условие}$   
 $\langle MV \rangle \rightarrow \langle MV \rangle | \langle SI \rangle$   
 $\langle E \rangle \rightarrow \langle E \rangle | \langle EI \rangle$   
 $\langle A \rangle \rightarrow \langle A \rangle | \langle AXP \rangle | \langle PI \rangle$

Вывод предложения с помощью данной грамматики означает, что слово, находящееся в предложении в позиции какого-либо терминала,

относится соответственно к действию, объекту или условию, в зависимости от типа терминала.

Отношения, найденные парсером, интерпретируются для представления технической функции согласно модели (1). Путем анализа определенных типов связей можно обнаружить слова, которые являются объектами действия, самим действием и условием выполнения действия (т.е. определить синтаксическую категорию).

Структуризация множества возможных значений компонента D «Действие» была произведена на основе набора терминов, наиболее часто встречающихся при описании функций технических объектов (например, в Международной классификации изобретений), а также выделенных посредством анализа неструктурированной информации в рубрике «Практическое применение» [14] имеющегося у авторов статьи фонда ФЭ (1200 описаний ФЭ) [15].

#### Классификатор действий (D) [2]:

- |                    |                     |                |
|--------------------|---------------------|----------------|
| • Преобразование   | • Соединение        | • Накопление   |
| • Изменение        | • Разделение        | • Выдача       |
| •• Увеличение      | • Стабилизация      | • Исследование |
| •• Уменьшение      | •• Параметрическая  | • Обработка    |
| • Измерение        | •• Пространственная | • Управление   |
| •• Прямое          | •• Временная        | • Контроль     |
| ••• Количественное | • Дестабилизация    | • Проведение   |
| ••• Качественное   | • Создание          | • Изолирование |
| •• Косвенное       | • Уничтожение       |                |

Использование данного классификатора позволяет найти общий класс для извлеченных действий из патентных формул. Общий класс будет являться признаком в морфологической таблице, а сами действия со своими объектами и условиями будут являться вариантами реализации (альтернативами) этого признака.

#### *Пример извлечения технических функций из патентов*

Текст патента: «Преобразователь сигнала, обеспечивающий формирование кодового слова цифровых данных в соответствии с методом построения кодового слова».



Рисунок 1 - Пример дерева разбора Link Grammar Parser

Из дерева зависимостей, показанного на Рисунке 1, согласно грамматике (2) получаем следующее:

- <W> →
  - обеспечивающий <MV> <E>
  - обеспечивающий формирование <MV> <E>
  - обеспечивающий формирование <A> слова <M> <E>
  - обеспечивающий формирование кодового слова <M> <E>
  - обеспечивающий формирование кодового слова <A> данных <E>
  - обеспечивающий формирование кодового слова цифровых данных <E>
  - обеспечивающий формирование кодового слова цифровых данных согласно
- <J>
  - обеспечивающий формирование кодового слова цифровых данных согласно методу <M>
  - обеспечивающий формирование кодового слова цифровых данных согласно методу построения <M>
  - обеспечивающий формирование кодового слова цифровых данных согласно методу построения <A> слова
  - обеспечивающий формирование кодового слова цифровых данных согласно методу построения кодового слова

Компоненты технической функции согласно разработанной модели (1), распознанные в тексте патентной формулы, будут следующие:

Действие (D) = обеспечивающий формирование

Объект (G) = кодового слова цифровых данных

Условие (H) = согласно методом построения кодового слова

### **Формирование морфологической таблицы**

Полученные компоненты представления технической функции согласно модели DGH (1), распознанные в сегментах предложений патентных формул, заносятся в таблицу технических функций вместе с выбранной обобщенной технической функцией (ТФ) согласно классификатору действий (D) (Таблица 1).

Таблица 1. Извлеченные технические функции

п/п	Действие (D)	Объект (G)	Условие (H)	Обобщенная ТФ
1	обеспечивает формирование	кодированное слово цифровых данных	согласно методу построения кодированного слова	Создание
2	выполняет	обратное быстрое преобразование Фурье	на постоперационных данных	Обработка
3	результат генерации	модулирующего сигнала	из ОБПФ вычислителя	Создание
4	формирует	давление	в пустой цилиндрической форме	Создание

На основе полученной таблицы, содержащей обобщенные технические функции, строится морфологическая таблица, в которой технические функции в виде кортежей «объект-условие-действие» DGH (1) используются в качестве альтернатив исполнения обобщенных технических функций. При построении морфологической таблицы (Таблица 2) просматриваются все найденные действия (D) и те действия, которые имеют одну и ту же обобщенную техническую функцию, записываются в один и тот же столбец морфологической таблицы со своими объектом и условиями.

Таблица 2. Морфологическая таблица

Создание	Обработка	...
D - обеспечивает формирование G - кодированное слово цифровых данных H - согласно методу построения кодированного слова	D - выполняет G - обратное быстрое преобразование Фурье H - на постоперационных данных	...
D- результат генерации G - модулирующего сигнала H- из ОБПФ вычислителя	...	...
D - формирует G - давление H - в пустой цилиндрической форме	...	...

Из полученной морфологической таблицы проектировщик выбирает интересующий его признак, т.е. обобщенную техническую функцию, и в столбце с соответствующим названием просматривает все возможные альтернативы ее исполнения. Обзор возможных способов реализации некоторой функции, сосредоточенных в одном месте, может подсказать



пользователю идею совершенствования имеющейся конструкции или создание совершенно нового технического решения.

### Результаты

Для хранения извлеченных технических функций и морфологической таблицы используется RDD (Resilient Distributed Dataset) схема для наиболее быстрой обработки данных о патентах за счет технологии MapReduce [16].

Для извлечения технических функций из патентных документов, создан программный модуль, состоящий из 5 блоков:

- блок обработки входного патентного документа;
- блок сегментации текста патентной формулы;
- блок семантического анализа текста;
- блок извлечения описаний технических функций;
- блок представления результатов обработки патентного массива.

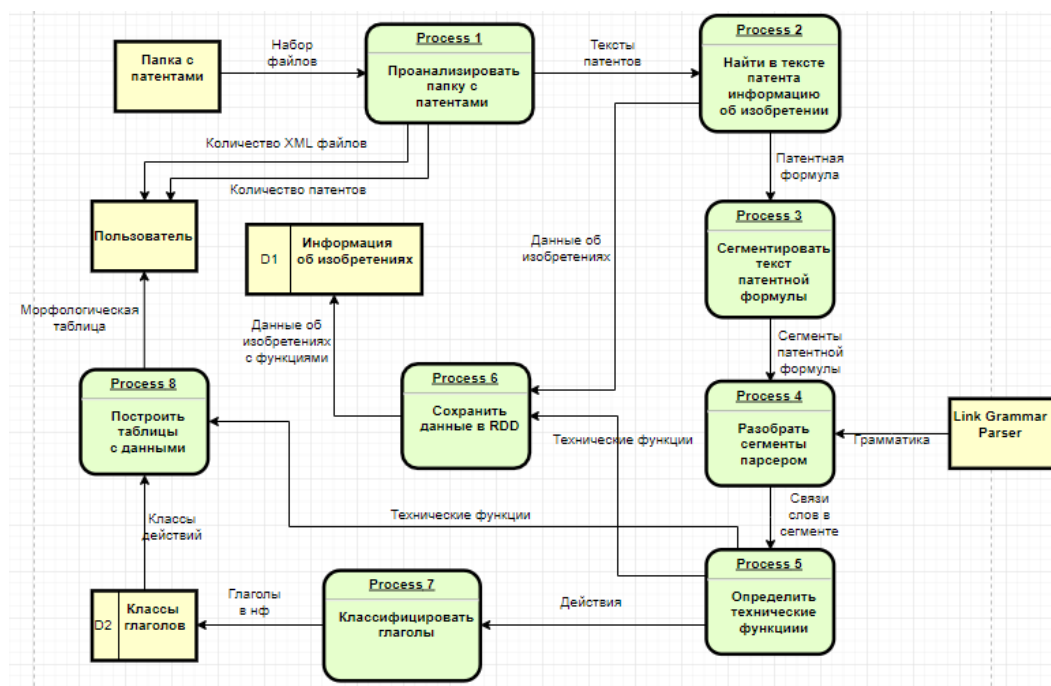


Рисунок 2 - Диаграмма потоков данных

### Обсуждение

Теоретическая ценность данной работы заключается: в разработанной грамматике представления описания технических функций в текстах русскоязычных патентов; в алгоритмах первичной обработки патентного массива, извлечения технических функций посредством анализа деревьев зависимостей, формирования морфологической таблицы.

Практическая значимость разработки заключается в реализованном программном модуле, состоящем из блоков извлечения информации из файла патента, блока семантического анализа, блока извлечения описания технической функции, блока формирования морфологической таблицы. Работоспособность программного модуля была проверена на ряде тестовых задач.

#### **Заключение**

*Работа выполнена при поддержке РФФИ (проекты №18-07-01086 А, № 16-07-00534 А).*

#### **ЛИТЕРАТУРА**

1. Половинкин А.И. Методы инженерного творчества: учеб. пособие / Половинкин А.И.; ВПИ. - Волгоград, 1984. - 365 с.
2. Фоменков С.А., Давыдов Д.А., Камаев В.А. Моделирование и автоматизированное использование структурированных физических знаний. Монография. - М.: Машиностроение - 1, 2004. - 278 с.
3. Давыдов Д.А., Фоменков С.А. Автоматизированное проектирование линейных структур физических принципов действия технических систем. //Машиностроитель, 2002, №2, с. 33-35.
4. Фоменков С.А., Колесников С.Г., Коробкин Д.М. Методика модификации информационного обеспечения базы данных физических эффектов. Известия Волгоградского государственного технического университета. 2012. № 10 (97). С. 166-170.
5. Дыков М.А., Кравец А.Г., Коробкин Д.М., Укустов С.М., Сальников М.Ю. Автоматизированная система принятия решений при патентной экспертизе. Известия Волгоградского государственного технического университета. 2014. № 6 (133). С. 35-41.
6. Коробкин Д.М., Фоменков С.А., Колесников С.Г. Автоматизация процесса формирования информационного обеспечения базы данных физических эффектов. Вестник компьютерных и информационных технологий. 2005. № 3 (9). С. 22-25.
7. Поисковая система ФИПС РосПатента [Электронный ресурс]. – 2018. – Режим доступа : [http://www1.fips.ru/wps/portal/IPS\\_Ru](http://www1.fips.ru/wps/portal/IPS_Ru) (дата обращ. 25.10.2018).

8. Кобликов И.А., Коробкин Д.М., Фоменков С.А., Яровенко В.А. Методика извлечения описаний реализуемых в патенте технических функций. Известия Волгоградского государственного технического университета. 2017. № 8 (203). С. 55-59.
9. Коробкин Д.М., Тюлькина Е.А., Фоменков С.А., Колесников С.Г. Система извлечения технических функций из патентного массива. ИТНОУ: Информационные технологии в науке, образовании и управлении. 2017. № 2 (2). С. 24-30.
10. Korobkin D., Fomenkov S., Kravets A., Kolesnikov S., Dykov M. Three-steps methodology for patents prior-art retrieval and structured physical knowledge extracting. Communications in Computer and Information Science. 2015. Т. 535. С. 124-136.
11. Link Grammar Parser [Электронный ресурс]. – 1998. – Режим доступа : <http://www.abisource.com/projects/linkgrammar> (дата обрац. 28.05.18).
12. Korobkin D.M., Fomenkov S.A., Kolesnikov S.G. A function-based patent analysis for support of technical solutions synthesis. В сборнике: 2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing, ICIEAM 2016 – Proceedings 2. 2016.
13. Хомский Н., Миллер Дж. Введение в формальный анализ естественных языков // Кибернетический сборник / Под ред. А.А.Ляпунова и О.Б.Лупанова. — М.: Мир, 1965.
14. Коробкин Д.М., Колесников С.Г., Фоменков С.А. Методика извлечения практического применения физических эффектов из патентных документов. В сборнике: Информационные технологии в науке, образовании и управлении материалы XLIV международной конференции и XIV международной конференции молодых учёных IT + S&E`16. под редакцией Е.Л. Глориозова. 2016. С. 92-100.
15. Korobkin D.M., Fomenkov S.A., Kolesnikov S.G. Ontology-Based extraction of Physical Effect description from Russian text. В сборнике: Proceedings of the European Conference on Data Mining 2014 and International Conferences on Intelligent Systems and Agents 2014 and Theory and Practice in Modern Computing 2014 - Part of the Multi Conference on Computer Science and Information Systems, MCCSIS 2014 2014. С. 260-262.
16. Изучаем Spark: молниеносный анализ данных / Х. Карау, Э. Конвински, П. Венделл, М. Захария. - Москва: ДМК Пресс, 2015. - 304 с.

S.S. Vasilyev, A.A. Kharitonov, D.M. Korobkin, S.A. Fomenkov  
**EXTRACTION OF MORPHOLOGICAL FEATURES OF TECHNICAL  
OBJECTS FROM RUSSIAN PATENTS**

*Volgograd State Technical University, Volgograd, Russia*

*The task of automation of synthesis of innovative solutions in the field of technical systems and technologies is one of the most priority problems of science. The authors propose to automate the most important, initial stages of the design of new technical systems and technologies on the basis of updated knowledge bases obtained from the world patent database, including the Rospatent patent database. According to the method of morphological analysis and synthesis, it is assumed that the main structural features (functions of technical objects) are extracted from some technical solution (patent), on the basis of which alternative options are sought. All these features are reduced to a morphological table, combined, which gives a lot of new solutions. The paper describes the developing a software for extracting the technical functions descriptions from Russian patents. The grammar of the presentation of technical functions descriptions according to the model "Action-Object-Condition" in the Russian-language patents was formed; algorithms for the initial processing of the patent database, the extraction of technical functions through the analysis of dependency trees, the formation of the morphological table was developed. The software consisting of a module of patent database processing; a module of text segmentation of the patent formula; a module of semantic text analysis; a module of extracting descriptions of technical functions; a module of presenting the results of patent database processing, was tested to solve practical problems.*

**Keywords:** technical functions, Natural Language processing, patents, RosPatent, Link Grammar Parser.

### REFERENCES

1. Polovinkin A.I. Metody inzhenerного tvorchestva: ucheb. posobie / Polovinkin A.I.; VPI. - Volgograd, 1984. - 365 s.
2. Fomenkov S.A., Davydov D.A., Kamaev V.A. Modelirovanie i avtomatizirovannoe ispol'zovanie strukturirovannyh fizicheskikh znaniy. Monografiya. - M.: Mashinostroenie - 1, 2004. - 278 s.
3. Davydov D.A., Fomenkov S.A. Avtomatizirovannoe proektirovanie linejnyh struktur fizicheskikh principov dejstviya tekhnicheskikh sistem. //Mashinostroitel', 2002, №2, s. 33-35.
4. Fomenkov S.A., Kolesnikov S.G., Korobkin D.M. Metodika modifikacii informacionnogo obespecheniya bazy dannyh fizicheskikh ehffektov. Izvestiya Volgogradskogo gosudarstvennogo tekhnicheskogo universiteta. 2012. № 10 (97). S. 166-170.
5. Dykov M.A., Kravec A.G., Korobkin D.M., Ukustov S.M., Sal'nikov M.YU. Avtomatizirovannaya sistema prinyatiya reshenij pri patentnoj ehkspertize. Izvestiya Volgogradskogo gosudarstvennogo tekhnicheskogo universiteta. 2014. № 6 (133). S. 35-41.

6. Korobkin D.M., Fomenkov S.A., Kolesnikov S.G. Avtomatizaciya processa formirovaniya informacionnogo obespecheniya bazy dannyh fizicheskikh ehffektov. Vestnik komp'yuternyh i informacionnyh tekhnologij. 2005. № 3 (9). S. 22-25.
7. Poiskovaya sistema FIPS RosPatenta. – 2018. – [http://www1.fips.ru/wps/portal/IPS\\_Ru](http://www1.fips.ru/wps/portal/IPS_Ru).
8. Koblikov I.A., Korobkin D.M., Fomenkov S.A., Yarovenko V.A. Metodika izvlecheniya opisaniy realizuemyh v patente tekhnicheskikh funkciy. Izvestiya Volgogradskogo gosudarstvennogo tekhnicheskogo universiteta. 2017. № 8 (203). S. 55-59.
9. Korobkin D.M., Tyul'kina E.A., Fomenkov S.A., Kolesnikov S.G. Sistema izvlecheniya tekhnicheskikh funkciy iz patentnogo massiva. ITNOU: Informacionnye tekhnologii v nauke, obrazovanii i upravlenii. 2017. № 2 (2). S. 24-30.
10. Korobkin D., Fomenkov S., Kravets A., Kolesnikov S., Dykov M. Three-steps methodology for patents prior-art retrieval and structured physical knowledge extracting. Communications in Computer and Information Science. 2015. T. 535. S. 124-136.
11. Link Grammar Parser. – 1998. <http://www.abisource.com/projects/linkgrammar>.
12. Korobkin D.M., Fomenkov S.A., Kolesnikov S.G. A function-based patent analysis for support of technical solutions synthesis. ICIEAM 2016
13. Homskij N., Miller Dzh. Vvedenie v formal'nyj analiz estestvennyh yazykov // Kiberneticheskij sbornik / Pod red. A.A.Lyapunova i O.B.Lupanova. — M.: Mir, 1965.
14. Korobkin D.M., Kolesnikov S.G., Fomenkov S.A. Metodika izvlecheniya prakticheskogo primeneniya fizicheskikh ehffektov iz patentnyh dokumentov. IT + S&E`2016. S. 92-100.
15. Korobkin D.M., Fomenkov S.A., Kolesnikov S.G. Ontology-Based extraction of Physical Effect description from Russian text. MCCSIS 2014. S. 260-262.
16. Izuchaem Spark: molnienosnyj analiz dannyh / H. Karau, EH. Konvinski, P. Vendell, M. Zahariya. - Moskva: DMK Press, 2015. - 304 s.