

В. Ф. Каблов, А. А. Рыбанов, Н. А. Кейбал, М. А. Маслова*, Д. А. Крюкова

**ПОВЫШЕНИЕ КАЧЕСТВА ДАННЫХ ПОСРЕДСТВОМ
СИНОНИМИЧЕСКОГО ПОДХОДА К ФОРМИРОВАНИЮ
ТЕЗАУРУСНОГО СЛОВАРЯ ИНГРЕДИЕНТОВ
И СВОЙСТВ РЕЗИНОВЫХ СМЕСЕЙ**

**Волжский политехнический институт (филиал)
Волгоградского государственного технического университета**

**E-mail: miss.mari.m@inbox.ru*

В работе описан разработанный программный модуль, обеспечивающий автоматизацию процессов управления данными в области эластомерных материалов. Ключевой особенностью модуля является использование синонимического подхода к формированию тезаурусного словаря ингредиентов и свойств резиновых смесей, что позволяет повысить качество и достоверность информационной базы. Полученная база данных может быть использована в дальнейших исследованиях для разработки и совершенствования резиновых смесей.

Ключевые слова: рецептура резин, база данных, тезаурус ингредиентов, программный модуль

Создание автоматизированных банков данных (АБнД) резиновых смесей позволяет существенно ускорить поиск необходимой техноло-

гической информации – рецептов с необходимым комплексом свойств, используемым ингредиентам, свойствам, технологическим ре-

жимам, областям применения резин и другой необходимой информации. АБнД являются необходимыми в разработке новых резиновых смесей с заданными свойствами. Создание АБнД требует решения ряда проблем, важнейшей из которых является создание эффективной, достаточной полной и логически выстроенной языковой подсистемы, описывающей предметную область базы данных. Актуальность проблемы языка в контексте развития информационных технологий является предметом пристального внимания разработчиков и предметных специалистов. Данная проблематика оказывает существенное влияние на построение и функционирование информационно-поисковых систем, баз данных и других технологических решений. Недостаточная лингвистическая компетентность зачастую приводит к получению нерелевантных результатов поиска и возникновению иных проблем. Решение языковой проблемы адекватного и однозначного описания терминов в информационной системе должна вестись в сотрудничестве программистов и предметных специалистов [1; 2].

Для упорядочения терминологии и однозначности описания в информационно-поисковых системах создаются указатели терминов, терминологические словари и классификаторы. ВИНТИ разработал указатель терминов по физике и физикохимии полимеров, а также терминологический справочник по резине. Перечень свойств позволяет вводить мнемонические сокращения наименований показателей для ускорения заполнения базы данных и уменьшения длины записей [3–6].

Разработка автоматизированной системы позволит осуществлять прогнозирование свойств эластомерного материала на начальных этапах проектирования по заданному составу ингредиентов. Кроме того, данная система будет способна выполнять обратное прогнозирование – определять состав материала, соответствующий заданным целевым свойствам. Такой подход обеспечит повышение эффективности процесса разработки рецептур резин, сократит временные и материальные затраты на экспериментальные исследования, а также позволит

оптимизировать состав материала в соответствии с требуемыми характеристиками.

Для разработки данной системы требуется создание комплексной справочной базы, включающей в себя подробные рецептурные описания, а также тезаурус, содержащий исчерпывающую информацию об используемых ингредиентах. Данный подход обеспечит необходимую информационную основу для эффективного функционирования системы. Тезаурус представляет собой множество смысловых единиц с заданной системой семантических отношений [2]. Применение тезаурусов и онтологий в промышленных информационных системах не распространено из-за сложности и недостаточного качества разрешения многозначности. Проблема формирования тезаурусов в области технологии резины отражена в работах [7; 8]. Подробно вопросы формирования тезаурусов в современных информационно-поисковых системах рассмотрены в [9].

Для формирования тезауруса ингредиентов резиновых смесей были проанализированы рецептурные карты различных заводов. На основе проведенного анализа были выработаны рекомендации по стандартизации названий. Эти рекомендации учитывали наиболее принятое употребление названий ингредиентов на основе товарных марок. Кроме того, в названиях, по возможности, исключались неинформативные термины. Например, название «сера молотая» имеет явную избыточность, так как комковая сера не употребляется в рецептуре. Поэтому для поискового предписания достаточно одного слова «сера». Какие-либо особенности ингредиентов приводятся в специальном примечании к рецепту (необходимость просева, возможные замены и т. п.).

Для различных групп ингредиентов были разработаны рекомендации по составлению тезауруса наименований.

В соответствии с принятыми правилами стандартизованные для поиска названия – дескрипторы (Д) записаны заглавными буквами, а остальные синонимы – строчными. В круглых скобках в необходимых случаях дается пояснение. Словарная статья в тезаурусе обычно имеет следующий вид:

ДЕСКРИПТОР
 (ПОЯСНЕНИЯ)
 с синонимы и условные синонимы
 в ВЫШЕСТОЯЩИЙ ДЕСКРИПТОР
 (ПОЯСНЕНИЯ)
 в ДР. ВЫШЕСТОЯЩИЕ ДЕСКРИПТОРЫ
 (ПОЯСНЕНИЯ)
 н НИЖЕСТОЯЩИЕ ДЕСКРИПТОРЫ
 а АССОЦИИРУЕМЫЕ ДЕСКРИПТОРЫ
 (ПОЯСНЕНИЯ)
 Например:
 НК
 с натуральный каучук
 светлый креп
 смокед-шитс
 в КАУЧУК
 КАУЧУК ИЗОПРЕНОВЫЙ
 ЭЛАСТОМЕР
 н НК ПЛАСТИКАТ
 а КАУЧУК СКИ-3

На представленной схеме отображена структурная диаграмма компонентов программного обеспечения, предназначенного для информационной поддержки в области свойств и рецептурного состава эластомерных материалов – базы данных по свойствам и составу рецеп-

тур эластомерных материалов.

В рамках модуля управления качеством данных, предназначенного для повышения качества ввода данных по ингредиентам и свойствам эластомерных материалов, разработаны три алгоритмических подхода:

Алгоритм 1. Алгоритм приведения марок ингредиентов к единому лексическому виду

Входные данные: Список выбранных марок ингредиентов $I = \{I(0), I(1), \dots, I(n)\}$

Выходные данные: единый лексический вид ингредиентов

```

for each I(i) where  $1 < i < n$  do
  if ингредиент I(i) в рецепте существует then
    заменяем ингредиент I(i) в рецепте на ингредиент I(0)
  удаляем ингредиент I(i)
  
```

Алгоритм 2. Алгоритм приведения свойств к единому лексическому виду

Входные данные: Список выбранных свойств $P = \{P(0), P(1), \dots, P(n)\}$

Выходные данные: Единый лексический вид свойств

```

for each P(i) where  $1 < i < n$  do
  if свойство P(i) в рецепте существует then
    заменяем свойство P(i) в рецепте на свойство P(0)
  удаляем свойство P(i)
  
```

Алгоритм 3. Алгоритм автоматического группирования марок ингредиентов

Входные данные: Общее наименование ингредиентов G и список его синонимов $S = \{S(0), S(1), \dots, S(n)\}$

Выходные данные: Общее наименование ингредиентов G и список марок ингредиентов $I = \{I(0), I(1), \dots, I(m)\}$

```

for each S(i) where  $0 < i < n$  do
  for each I(j) where  $0 < j < m$  do
    if наименование ингредиента I(j) содержит наименование синонима S(i) then
      заменяем общее наименования ингредиента I(j) на общее наименование G
  
```

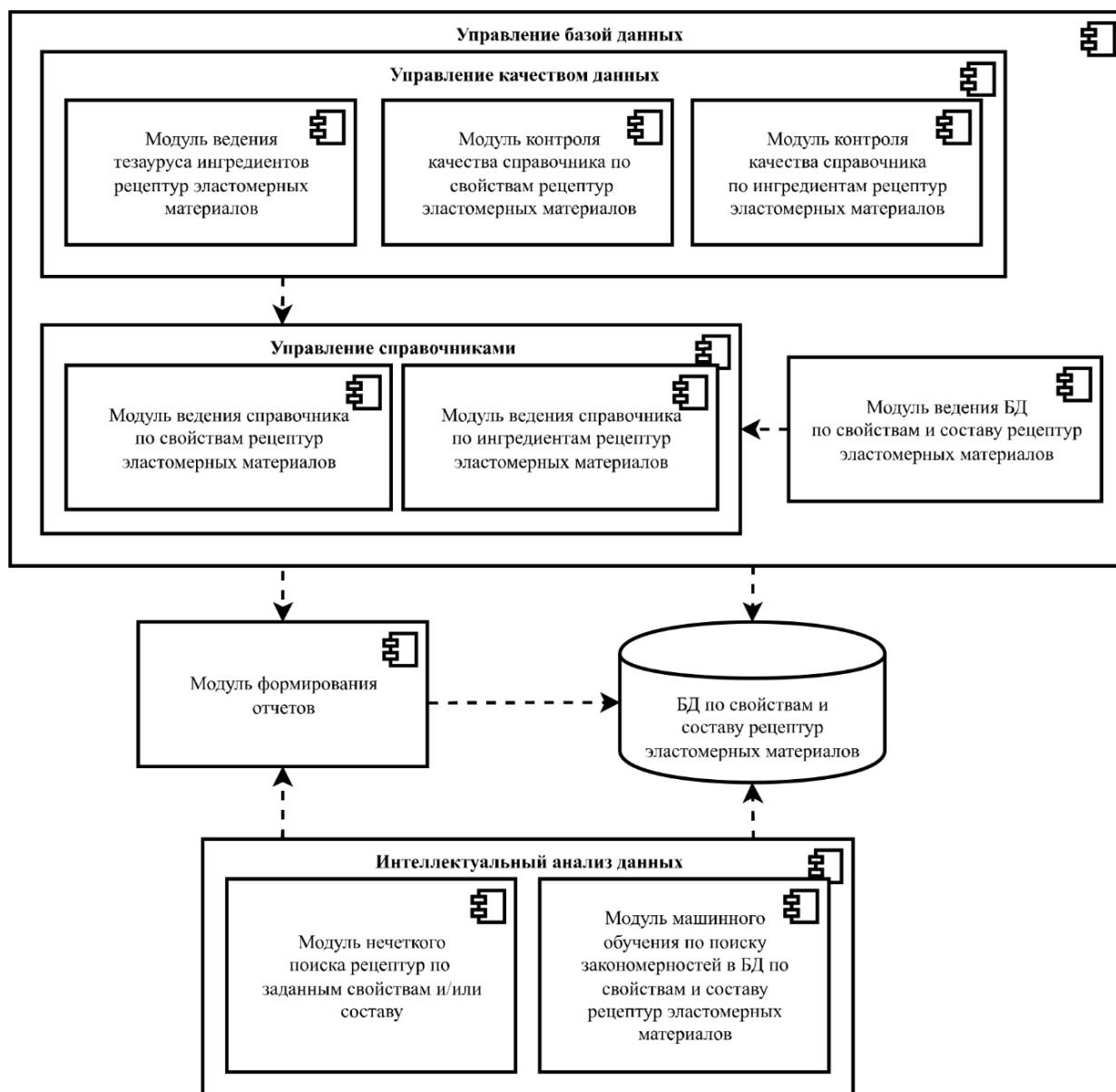


Диаграмма компонентов программы для информационной поддержки в области свойств и рецептурного состава эластомерных материалов

На основе указанных алгоритмов была разработана программа повышения качества ввода данных в базу данных. Программа интегрирована с автоматизированным банком данных по рецептурам резиновых смесей «Rubber C» [10; 11].

Таким образом, предлагаемый программный модуль повышает качество данных в резиновых смесях с помощью синонимического подхода. Тезаурусный словарь ингредиентов и свойств резиновых смесей формируется на основе семантических связей между терминами. Учет семантических связей позволяет расши-

рить и структурировать информационное наполнение базы данных. Повышение качества и достоверности информации способствует анализу и моделированию эластомерных материалов. Применение разработанного программного модуля позволяет более эффективно прогнозировать свойства эластомерных материалов и оптимизировать их рецептурный состав. Накопленные данные могут служить основой для теоретических и экспериментальных исследований в области разработки и совершенствования резиновых смесей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Каблов, В. Ф. Информационные технологии в разработке и в производстве эластомерных материалов / В. Ф. Каблов, И. М. Агаянц. – Волгоград, 2009. – 409 с.
2. Каблов, В. Ф. Применение информационных и цифровых технологий в разработке и исследовании эластомерных материалов: монография / В. Ф. Каблов; ВПИ (филиал) ВолгГТУ. – Волгоград, 2021. – 364 с.
3. Аскадский, А. А. Указатель терминов по физике и физико-химии полимеров / А. А. Аскадский, Т. Л. Ренард, И. М. Пашков. – М.: ВИНТИ, 1983. – 45 с.
4. Махлис, Ф. А. Терминологический справочник по резине / Ф. А. Махлис, Д. Л. Федюкин. – М.: Химия, 1989. – 400 с.
5. Солтон, Дж. Динамические библиотечно-информационные системы / Дж. Солтон. – М.: Мир, 1979. – 558 с.
6. Шведков, Е. Л. Тезаурус информационно-поисковый по порошковой металлургии / Е. Л. Шведков, И. И. Ковенский. – Киев: Наукова думка. – 1984. – 360 с.
7. Каблов, В. Ф. Использование автоматизированных банков данных для проектирования рецептур резин / В. Ф. Каблов, В. П. Шевчук. – Волгоград: ВолгПИ. – 1988. – 112 с.
8. Каблов, В. Ф. Технологическая информатика полимерных материалов / В. Ф. Каблов. – Волгоград: ВолгПИ, 1990. – 112 с.
9. Лукашевич, Н. В. Тезаурусы в задачах информационного поиска / Н. В. Лукашевич. – М., 2010. – 396 с., ил
10. Каблов, В. Ф. Автоматизированный банк данных нового поколения рецептур и свойств резин - «Раббер С» / В. Ф. Каблов, А. А. Рыбанов, Н. А. Кейбал // «Каучук и резина» 2024, № 3, с. 168–173.
11. Свид. о гос. регистрации программы для ЭВМ № 2024622270 от 24 мая 2024 г. Российская Федерация. База данных по рецептурам, свойствам и группам резиновых смесей / В. Ф. Каблов, А. А. Рыбанов, Н. А. Кейбал; ФГБОУ ВО ВолгГТУ. – 2024.

REFERENCES

1. Kablov, V.F., Agayancz I.M. Informacionny`e texnologii v razrabotke i v proizvodstve e`lastomerny`x materialov. – Volgograd, 2009. – 409s.
2. Kablov, V.F. Primenenie informacionny`x i cifrovyy`x texnologij v razrabotke i issledovanii e`lastomerny`x materialov: mono-grafiya / V.F. Kablov; VPI (filial) Volg-GTU. - Volgograd, 2021. - 364 s.
3. Ukazatel` terminov po fizike i fiziko-ximii polimerov/ Askadskij A.A., Renard T.L., Pashkov I.M..- M.:VINITI, 1983.-45s.
4. Maxlis, F.A., Fedyukin D.L. Terminologicheskij spravochnik po re-zine. –M.: Ximiya, 1989.-400s.
5. Solton, Dzh. Dinamicheskie bibliotечно-informacionny`e sistemy`. - M.: Mir. – 1979. - 558 s.
6. Shvedkov, E.L. Kovenskij I.I. Tezaurus informacionno-poiskovy`j po poroshkovej metallurgii. - Kiev: Naukova dumka. - 1984.-360 s.
7. Kablov, V.F., Shevchuk V.P. Ispol`zovanie avtomatizirovanny`x bankov danny`x dlya proektirovaniya receptur rezin.- Volgograd: VolgPI. – 1988. -112s.
8. Kablov, V.F. Texnologicheskaya informatika polimerny`x materia-lov. Volgograd: VolgPI. 1990.- 112s.
9. Lukashevich, N. V. Tezaurusy` v zadachax informacionnogo poiska – M., 2010. – 396 s., il
10. Kablov, V.F. Ry`banov A.A., Kejbal N.A. Avtomatizirovanny`j bank danny`x novogo pokoleniya receptur i svojstv rezin - «Rabber S» // «Kauchuk i rezina» 2024, №3, s.168-173.
11. Svid. o gos. registracii programmy` dlya E`VM № 2024622270 ot 24 maya 2024 g. Rossijskaya Federaciya. Baza danny`x po recepturam, svojstvam i gruppam rezinovy`x smesey / V.F. Kablov, A.A. Ry`banov, N.A. Kejbal; FGBOU VO VolgGTU. - 2024.

V. F. Kablov, A. A. Rybanov, N. A. Keibal, M. A. Maslova, D. A. Kryukova
IMPROVE DATA QUALITY THROUGH A SYNONYMIC APPROACH
TO BUILDING A THESAURUS DICTIONARY
OF RUBBER COMPOUNDS INGREDIENTS AND PROPERTIES

Volzhsky Polytechnic Institute (branch)
Volgograd State Technical University

Abstract. The paper describes the developed software module providing automation of data management processes in the field of elastomeric materials. The key feature of the module is the use of synonymic approach to the formation of thesaurus dictionary of ingredients and properties of rubber mixtures, which allows to increase the quality and reliability of the information base. The obtained database can be used in further research for the development and improvement of rubber mixtures.

Keywords: rubber, database, ingredients thesaurus, program module