



В. С. Крук, З. А. Антонова, Г. Я. Кабо,
О. В. Говин, Ю. В. Максимук

ИНФОРМАЦИОННЫЕ БАЗЫ ДАННЫХ ДЛЯ ХИМИЧЕСКОЙ ПРОМЫШЛЕННОСТИ РЕСПУБЛИКИ БЕЛАРУСЬ

Необходимость совершенствования технологии производств органического синтеза и нефтепереработки обусловлена проблемами экологии, энерго- и ресурсосбережения. Доля затрат на сырье и энергию в промышленном органическом синтезе в Республике Беларусь на 30–60 % выше, чем в аналогичных зарубежных производствах. Энергозатраты на основных производствах азотной промышленности выше примерно на 35 %. Частичное решение этих проблем может быть достигнуто в результате сокращения числа побочных продуктов, являющихся отходами производства, и за счет использования их в качестве сырья для малотоннажных производств [1–3].

Темпы научно-технического прогресса и интенсификации промышленности зависят главным образом от развития эффективных методов и средств управления наукой и техникой. Принятие своевременных решений о проведении исследований стало невозможным без глубокого анализа текущей и прогнозной информации, которую в современных условиях можно обработать только с применением сетей ЭВМ, персональных компьютеров, а также информационных баз данных.

Поиск информации и доступ к ней специалистов и ученых на различных этапах выполнения научно-исследовательских и опытно-конструкторских работ превратился в отдельную проблему. Отсутствие необходимых сведений о мировых достижениях и тенденциях развития науки и техники, а также несоответствие системы научно-технической, экономической и коммерческой информации требованиям сегодняшнего дня являются одной из причин отставания нашей страны по уровню выпускаемой продукции. Другая причина состоит в существенных различиях принципов развития промышленности органического синтеза Республики Беларусь и промышленно развитых стран. В республике предпочтение было отдано строительству многотоннажных производств органического синтеза, почти полностью работающих на привозном сырье. В настоящее время в республике работает 67 предприятий химической и нефтехимической промышленности. Мощность производств основных продуктов составляет (в тыс. т/год): капролактам – 105,8, диметилтерефталат – 187,5, азотные удобрения – 607,7, бензол – 40,2, ксилол – 27,8 и акрилонитрил – 62,7 [4]. Зарубежный опыт свидетельствует о предпочтительности организации гибких производств с очень широкой номенклатурой химических ве-

ществ и товаров. Существенно более широкий ассортимент веществ и товаров делает производство экономически более стабильным, конкурентоспособным на внутреннем и мировом рынках, позволяет достигать высокой экономической эффективности использования сырья, энергии и уменьшения нагрузки на экологические системы. Поэтому главным направлением развития промышленности органического синтеза в Беларуси должно стать создание на основе комплексного использования сырья небольших производств высокотехнологических продуктов с высокой стоимостью на мировом рынке.

Расширение номенклатуры химических веществ и товаров на действующих производствах не может быть произвольным. Оно должно обосновываться логикой возможных превращений некоторых базовых компонентов и потребительскими качествами веществ и товаров, изготавливаемых из них. Фундаментом для этого служат информационные базы данных, аккумулирующие сведения о способах производства химически родственных групп веществ, их физико-химических свойствах, потребительских характеристиках, альтернативных технологиях, катализаторах процессов и т. п., а также блоки расчетных программ для определения термодинамически обоснованных условий синтеза, параметров, необходимых для проектирования, анализа веществ и т. п.

В лаборатории термодинамики органических веществ и на кафедре физической химии БГУ на протяжении ряда лет ведутся работы по созданию специализированных баз научно-технических данных. Выполнено обоснование принципов рациональной структуры информационно-аналитических систем для отраслей промышленности органического синтеза (капролактамы – «Lactam», мочевины и ее производные – «Карбамид», полиэтилентерефталат – «Лавсан») и переработки нефти и ее фракций («Стандартные химические эксергии С, Н, О, N, S-содержащих органических веществ»). Разработаны методики поиска, принципы анализа и формализации информационных источников, создано программное обеспечение для их функционирования. Основными источниками информации служили публикации в различных научных и научно-производственных журналах, описания патентов, отчеты по НИР и НИОКР, материалы конференций, диссертаций.

В последнее время существенную помощь в получении информации оказывают различные электронные базы данных, доступные через Internet, в частности, ВИНТИ предлагает электронную версию реферативного журнала «Химия» [5]. Кроме библиографических баз данных имеются специализированные электронные системы по свойствам веществ. Эти системы содержат информацию, полученную на основе обработки большого числа литературных источников. Например, базы данных по термодинамическим свойствам веществ ИВТАНТЕРМО (Москва) и TRC (США).

База ИВТАНТЕРМО представляет собой автоматизированную систему, предназначенную для обеспечения научных работников, инженеров и конструкторов информацией о термодинамических свойствах индивидуальных веществ в стандартном состоянии в широком интервале температур. Для всех веществ, содержащихся в базе ИВТАНТЕРМО, хранится следующая термодинамическая и вспомогательная информация: химическая формула, название вещества, молекулярная масса, энтальпия образования при 298,15 К, теплоемкость (C_p) при 298,15 К, энтропия при 298,15 К, теплоемкость в широком температурном интервале (для газов).

База TRC создавалась Центром термодинамических исследований Техасского университета (с 1961 г.), который занимается изучением, оценкой, отбором и публикацией наиболее достоверных экспериментально полученных данных о физических и термодинамических свойствах химических соединений и их смесей. В настоящее время Термодинамический исследовательский центр национального института стандартов и технологии создает глобальную систему физико-химической информации Researcher. Она включает в себя хранилище информации, доступное на бесплатной основе, в том числе и для белорусских организаций и пользователей [6], стандарты [7], программное обеспечение для сбора и анализа информации (GDC). Информация включает теплофизические и термохимические характеристики веществ, смесей и химических реакций, необходимые для научных исследований и создания новых материалов, разработки химических процессов, энергетики, экологии и санитарии. В настоящее время действует компьютерная программа для сбора информации и договор с издательством научного журнала «Journal of Chemical and Engineering Data» о том, что публикуемая в нем информация обрабатывается GDC и помещается в общедоступное хранилище. Функционирующая система позволяет решить ряд задач по сохранению информации, ее доступности, скорости получения, автоматическому контролю качества данных, что ведет к высвобождению человеческих ресурсов при обработке данных, развитию «компьютерной науки», когда компьютеры выявляют закономерности и аномалии.

Кроме этого имеется большое число специализированных баз данных по свойствам отдельных классов соединений, их синтезу, методам исследований и контролю качества продукции, автоматизации производства и т. д. [8–12].

Анализ различного рода электронных баз позволил сформулировать основные требования к ним. База должна обеспечивать более высокий, чем обычные универсальные базы, уровень концентрации информации, более удобную структуру распределения информации, существенно большее разнообразие по объему (рефераты, тексты статей, патентов, отчетов) и виду информации (тексты, таблицы, графики, схемы и т. п.), возможность использования вспомогательных программ для переработки информации из основной базы данных (например, программы расчета термодинамических свойств, химических и фазовых равновесий, химических эксергий и эксергетических балансов).

ОТРАСЛЕВАЯ ИНФОРМАЦИОННАЯ БАЗА ДАННЫХ «LACTAM»

В качестве образца информационно-аналитической системы нами была создана специализированная база научно-технических данных для одного из крупнейших производств промышленности органического синтеза Республики Беларусь – капролактама [13].

В основу базы «Lactam» положена иерархическая структура данных на основе ключевых слов-признаков, образованных из названий используемых веществ, технологических операций, свойств соединений и т. п. (рис. 1). Она представляет собой строго ограниченный набор из 89 ключевых слов-признаков и неограниченного дополнительного списка по побочным и вспомогательным веществам. Информация систематизируется на основе древовидной раз-

ветвящейся структуры с использованием ключевых слов различного уровня. Ключевые слова 1-го уровня образуют названия исходных, промежуточных и конечных продуктов производства капролактама. Следующий уровень образуют слова, дающие общие характеристики соединений (свойства веществ, химические реакции, методы анализа, технология и экономика производства, экология и охрана труда). Дальнейшее повышение уровня ключевого слова приводит к конкретизации понятия. Например, общий раздел «Свойства вещества» включает подразделы: физико-химические свойства, термодинамические, транспортные, оптические и т. д. Такой подход к разграничению ключевых слов по уровням позволил создать достаточно ясную и универсальную структуру базы. В литературных источниках кроме основных соединений дается описание и других встречающихся в тексте веществ, их свойств, применения (например, продукты реакции с участием основных соединений, побочные вещества технологических процессов и т. п.). Эти соединения также вводятся в базу данных, но с меньшей степенью детализации, образуя отдельный раздел «Дополнительные вещества».

Анализ содержания информационного источника выполняется экспертом, который определяет уровень значимости информации и принимает решение о его дальнейшей обработке. Обычно первичный анализ информации заключается в отборе всех данных, относящихся к областям исследований. Далее на основании отобранных ключевых слов производится описание источника информации. Причем описание может производиться достаточно формально на основании проверки на присутствие определенной комбинации ключевых слов в тексте. Но не всегда наличие реперных слов (это, как правило, касается ключевых слов первого, реже – второго порядка) является критерием выбора адреса информационного источника в базе данных. Иногда трудно разграничить информацию соответственно степени ее значимости. Только квалификация эксперта позволяет сделать правильный вывод о соответствии источника теме исследования и оценить полезность этих данных для НИР.

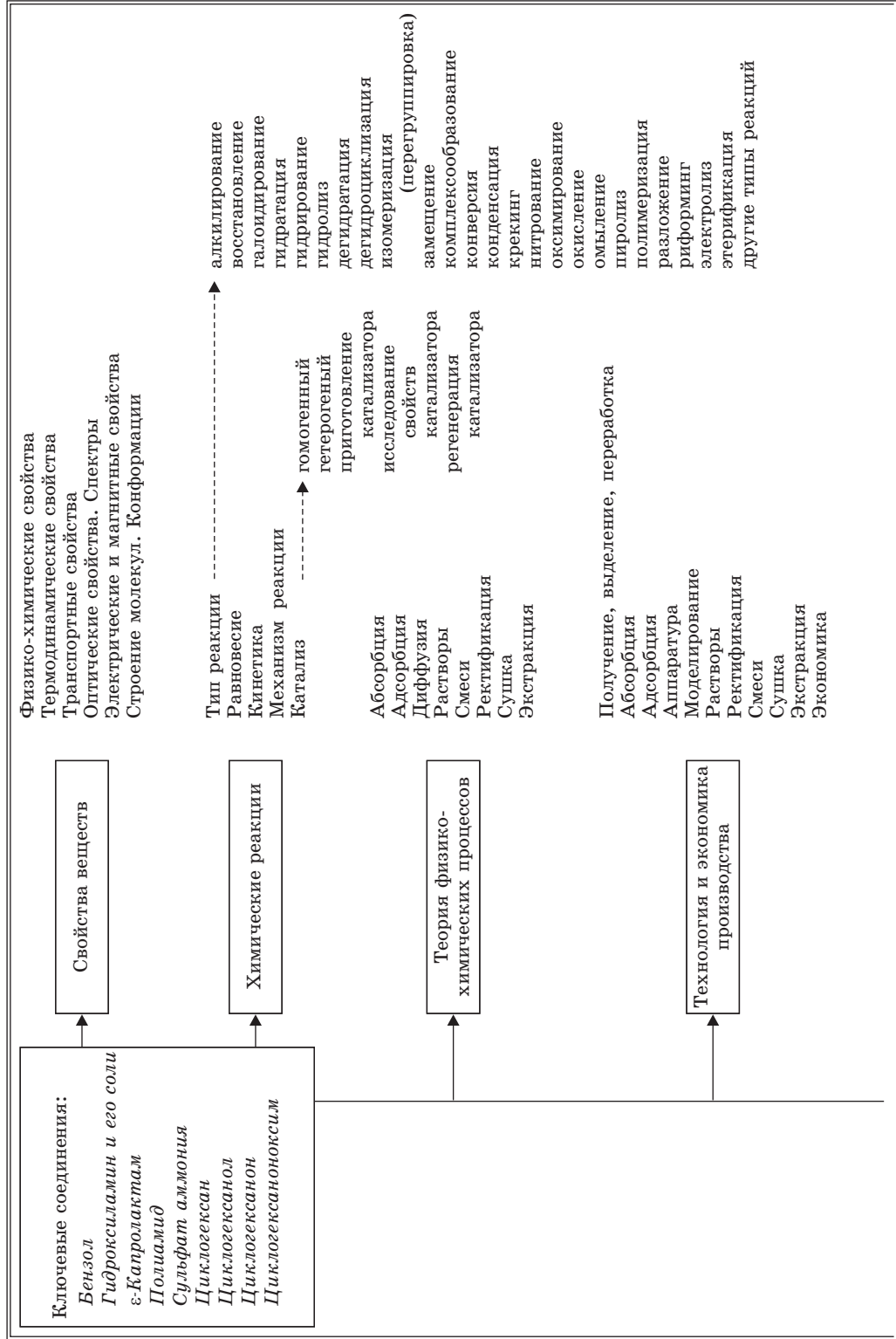
Система управления базой данных разработана на основе системы программирования баз данных FoxPro 2.6 для Windows [10, 14]. Основная программа состоит из трех подпрограмм:

- 1) ввода и регистрации текста;
- 2) осуществления поиска по ключевым словам и авторам;
- 3) редактирования структуры базы данных.

При своей работе подпрограммы используют следующие справочники и таблицы данных:

- 1) справочник ключевых слов;
- 2) справочник связи текст – автор;
- 3) таблицу описания текстов по ключевым словам;
- 4) текстовые файлы рефератов.

Предлагаемая программа позволяет проводить наполнение базы данных в режиме диалога, хранить информацию, выполнять поиск необходимых сведений и редактировать список ключевых слов, причем объем хранимых данных определяется лишь техническими характеристиками используемой ЭВМ. Она поддерживает все режимы работы с операционной системой Windows 95/98 и имеет современный дизайн оформления.



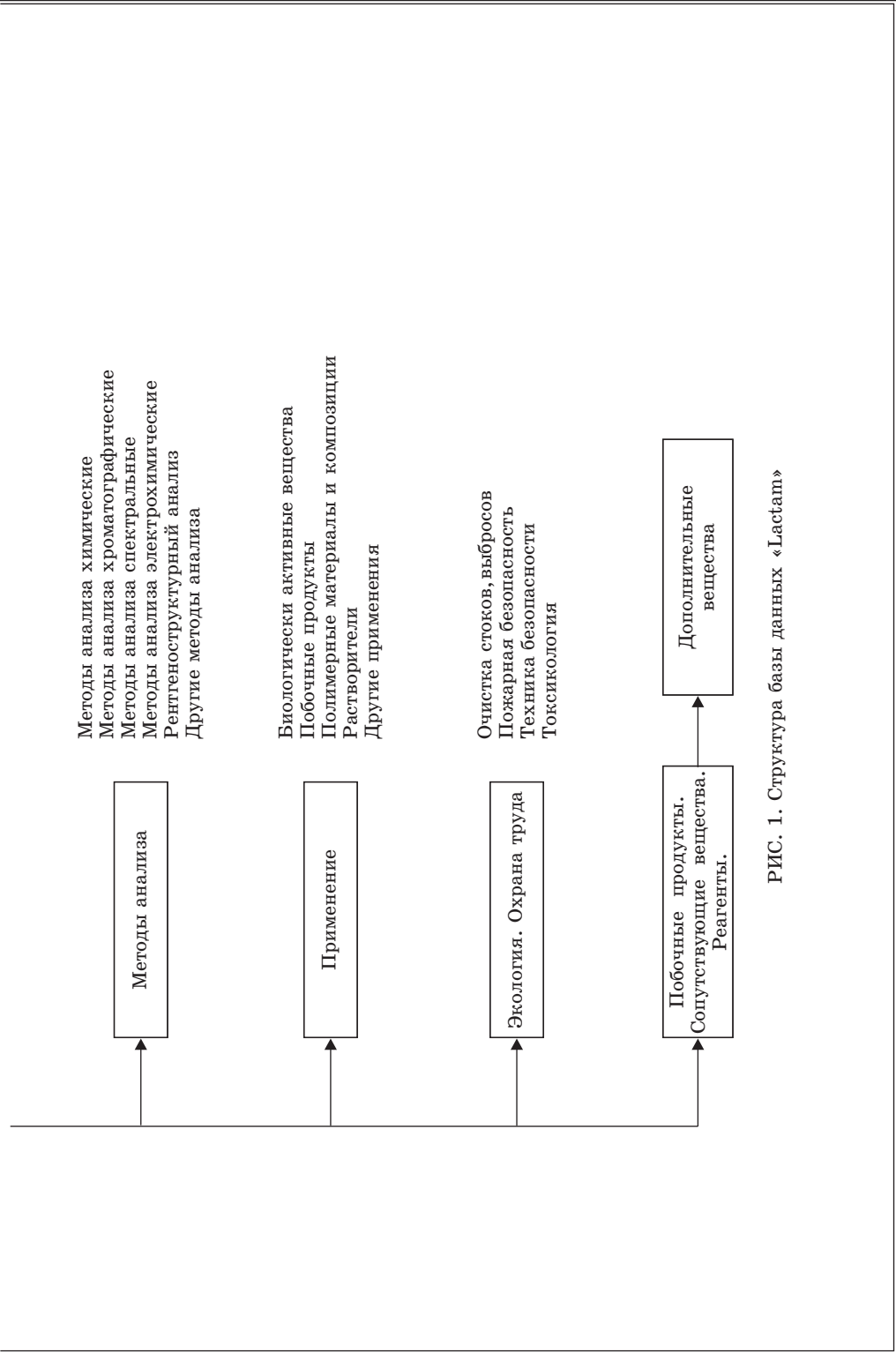


РИС. 1. Структура базы данных «Lastam»

Для пользователей основным режимом работы базы данных является поиск информации. Он может осуществляться по ключевым словам или авторам. В результате работы подпрограммы поиска выводится список рефератов статей и их авторов по заданному запросу, с текстами которых можно ознакомиться, обратившись к встроенной подпрограмме текстового редактора. Далее текст реферата может быть при необходимости распечатан или дополнительно отредактирован для вставки в другой документ.

Работа с базой данных в режиме «Поиск» осуществляется последовательным выбором ключевых слов разных уровней. Например, для нахождения физико-химических свойств циклогексана последовательно отмечаются слова-признаки: «Циклогексан» → «Свойства веществ» → «Физико-химические свойства». После этого включается программа поиска реферата. В результате открывается окно с таблицей (рис. 2). В статусной строке указано, какое количество рефератов найдено. В таблице предусмотрено поле для хранения логического пути текстов рефератов на жестком диске, например «C:\BASE-II\1989\6b2624.TXT», авторы этой статьи: «Sun T. F., Schouten J. A., Traprenniers N. J., Biswas J. A.». В дополнительном окне показаны код реферата по РЖХ, название статьи, авторы, выходные данные журнала, в

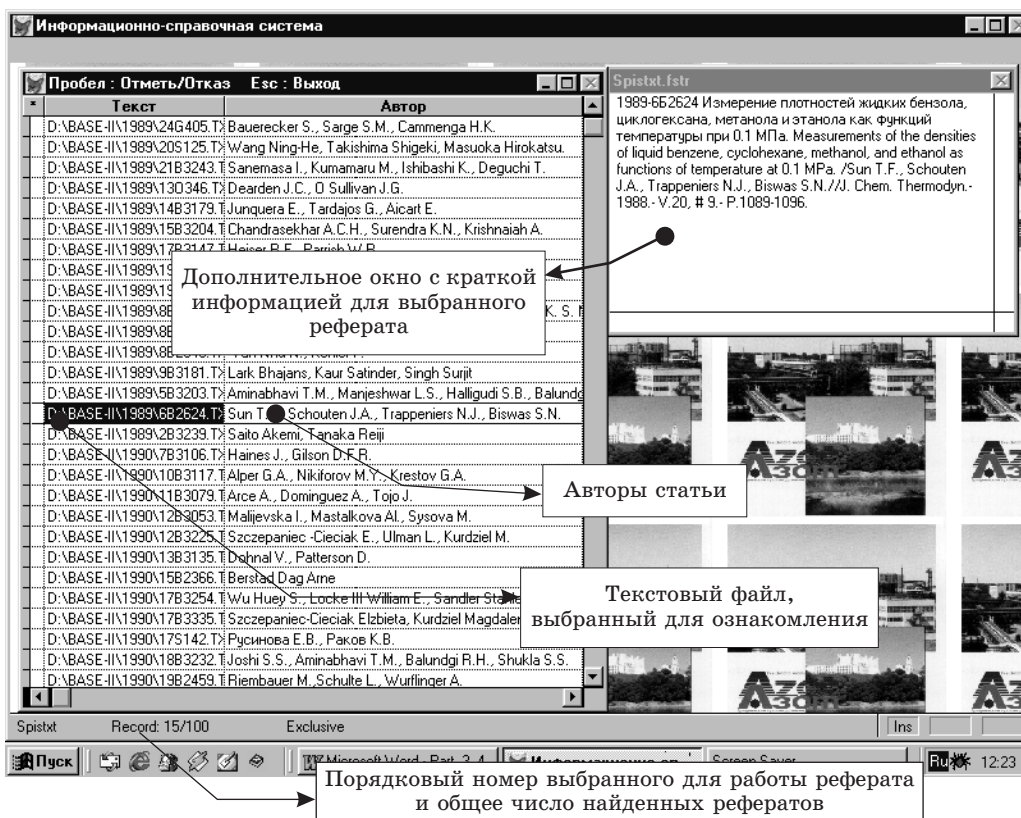


РИС. 2. Результаты поиска по физико-химическим свойствам циклогексана для базы данных «Lactam»

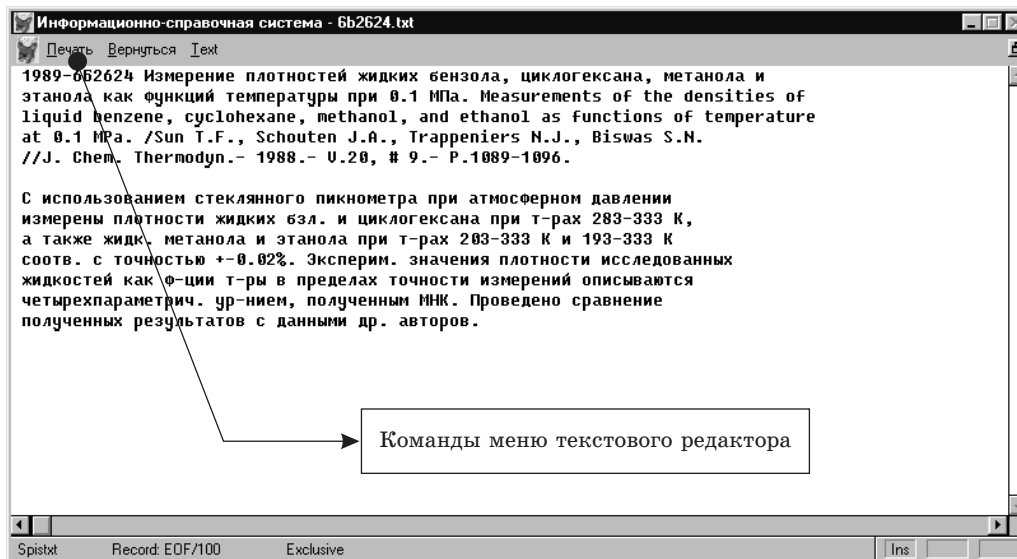


РИС. 3. Окно текстового редактора с текстом выбранного реферата 6b2624.TXT

котором опубликована статья. Если пользователю необходимо ознакомиться с текстом реферата, его нужно отметить и запустить подпрограмму текстового редактора. В новом окне текст реферата может быть распечатан, дополнительно отредактирован для вставки в другой документ и сохранен на жестком диске (рис. 3).

Чтобы найти рефераты по всем физико-химическим, термодинамическим, транспортным и другим свойствам циклогексана, следует выбрать раздел «Свойства веществ» второго подуровня. При этом будут найдены все рефераты, имеющие ключевые слова более низкого уровня.

Третья команда пункта «Поиск» – «Поиск по авторам». На экране разворачивается таблица, в которой указаны имя текстового файла и авторы статьи. Далее поиск по авторам можно проводить в ручном либо автоматическом режиме. Автоматический поиск может проводиться также и по части фамилии автора. Фамилия автора, введенная разными алфавитами (латинским или кириллицей), воспринимается программой как две различные.

Реферативная база данных «Lactam» предназначена для многоаспектного поиска опубликованных работ по технологии синтеза капролактама и сопутствующих веществ. База данных позволяет получить информацию о примерно 4 000 литературных источниках для основных соединений производства капролактама, о химико-технологических процессах с их участием, а также способах дополнительной переработки этих соединений.

Информационно-аналитическая система «Lactam» кроме своего основного назначения – поиска литературных источников – может использоваться для статистического анализа информации по производству капролактама.

Для этого определяем общее распределение числа литературных источников по основным веществам, участвующим в технологическом процессе производства капролактама. Чтобы найти общее число ссылок по свойствам, реакциям, применению бензола, следует в режиме «Поиск» отметить только слово

«Бензол». После завершения работы подпрограммы поиска в нижней части открывшегося окна указано, что найдено 1800 источников. Повторив указанную процедуру для остальных веществ, получим распределение числа литературных источников по основным веществам, как представлено на рис. 4.

Общее число источников, которые появляются в ответ на запрос, – 5127 ссылок, хотя в базе данных описано 3764 реферата. Такое несоответствие объяснимо. В большинстве рефератов встречается одновременно несколько ключевых слов. Например, если в реферате описывается реакция изомеризации циклогексаноноксида в капролактамы, то ссылка на него будет дана дважды: на слова «Капролактамы» и «Циклогексаноноксид». Кроме этого, все побочные вещества описываются параллельно, по крайней мере, с одним из основных веществ.

Как видно из рис. 4, наибольший объем информации представлен по свойствам бензола. Сравнительно много информации по циклогексану и побочным продуктам.

Ниже приведены данные по распределению литературных источников по свойствам циклогексана (рис. 5). Процедура определения числа ссылок такая же, как была описана для бензола, только выбираются конкретные свойства циклогексана.

Как следует из рис. 5, в информационных источниках очень мало внимания уделяется вопросам применения циклогексана и экологическим проблемам.

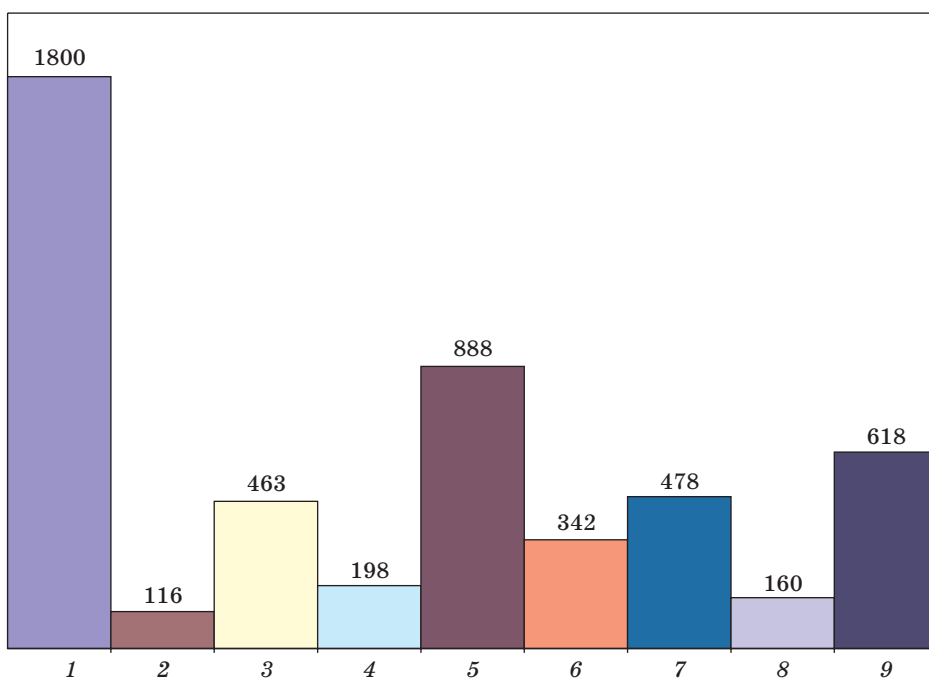


РИС. 4. Распределение числа литературных источников по веществам:
1 – бензол, 2 – гидроксилламин и его соли, 3 – ε-капролактамы, 4 – сульфат аммония,
5 – циклогексан, 6 – циклогексанол, 7 – циклогексанон, 8 – циклогексаноноксид,
9 – дополнительные вещества

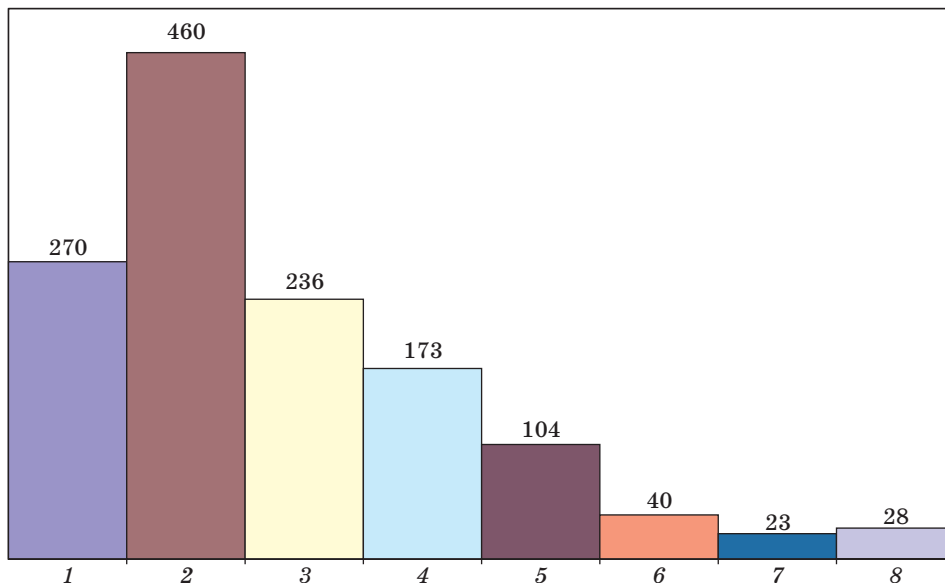


РИС. 5. Распределение числа литературных источников в базе данных «Lactam» для циклогексана по его свойствам:

1 – свойства вещества, 2 – химические реакции, 3 – теория физико-химических процессов, 4 – технология и экономика производства, 5 – методы анализа, 6 – применение, 7 – побочные продукты, 8 – экология

В заключение приведем пример распределения литературных ссылок по методам анализа основных веществ. По сравнению с первыми двумя примерами в этом случае демонстрируется третья ступень детализации понятия, которая касается уже конкретных методов анализа. Распределение числа литературных источников по методам анализа приведено на рис. 6.

Следует отметить, что на этом рисунке приведено общее число источников по каждому из методов анализа для всех соединений. Для этого в режиме «Поиск», например, ключевое слово «Спектральные методы анализа» отмечалось на каждой из ветвей основных соединений структуры базы данных. Данные диаграммы свидетельствуют о том, что значительная доля информации попадает в общий раздел «Другие методы анализа», не детализуясь. В этот раздел также будут попадать и все новые методы анализа, но это не столь существенно для ретроспективных баз данных, какой и является база данных «Lactam».

Функциональные возможности любой реферативной базы данных легко могут быть расширены. Например, база данных «Lactam» по мере замены текстов рефератов статей на оригинальные работы превращается в полнофункциональный библиографический справочник по производству капролактама. Другое направление модификации реферативной базы данных выражается в наполнении ее значимыми данными из оригинальных работ. При этом такая база данных может стать физико-химическим и термодинамическим справочником при сравнительно небольших затратах времени. Так, в реферативную базу данных «Lactam» был введен блок по термодинамическим и физико-химическим свойствам основных соединений, участвующих в производстве кап-

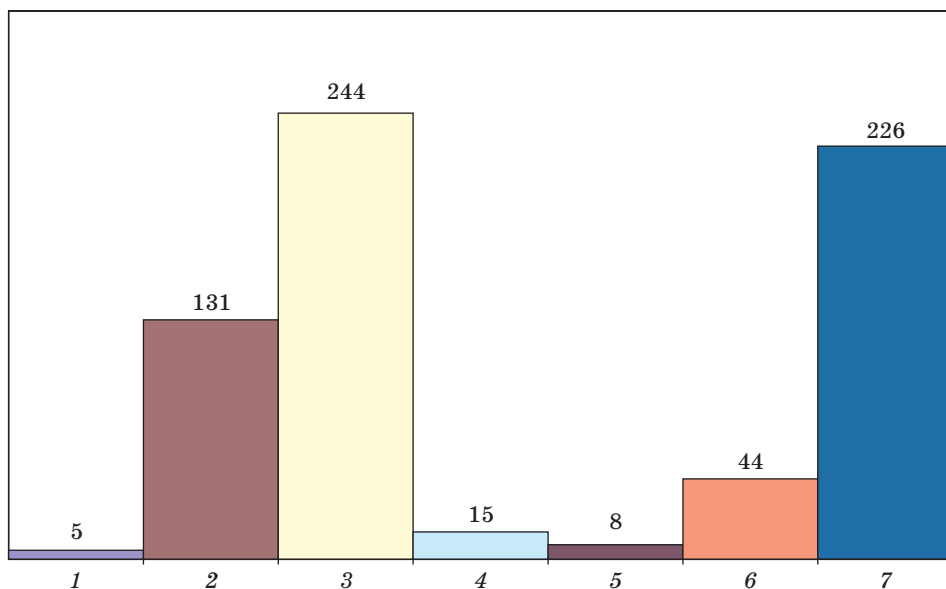


РИС. 6. Распределение числа литературных источников в базе данных «Lactam» по различным методам анализа:

1 – химические, 2 – хроматографические, 3 – спектральные, 4 – электрохимические, 5 – термический, 6 – рентгеноструктурный, 7 – другие

ролактама. В соответствующие разделы «Свойства веществ» – «Физико-химические свойства» и «Термодинамические свойства» были добавлены подуровни «Таблицы», в которых приведена информация по физико-химическим, термодинамическим свойствам веществ из оригинальных работ, тексты рефератов которых представлены в базе данных. Каждая из таблиц по возможности содержит сведения по названию соединения, его молекулярной массе, плотности, температурам плавления и кипения, давлению насыщенных паров при различных температурах, поверхностному натяжению, показателям преломления, вязкости, теплоемкостям, энтальпиям и энтропиям фазовых превращений, энергии сгорания и образования, энтропии.

Описанная выше разработка была использована при создании специализированной базы данных «Лавсан» по физико-химическим свойствам, методам синтеза и применению толуиловых кислот (орто-, мета-, пара-) и их метиловых эфиров для могилевского ПО «Химволокно».

Паратолуиловая кислота и параметилтолуилат являются промежуточными веществами и основными полупродуктами в синтезе терефталевой кислоты и диметилтерефталата, из которых в конечном итоге синтезируется полиэтилентерефталат – один из наиболее распространенных полимерных продуктов. Орто- и метатолуиловые кислоты и их метиловые эфиры образуются в указанных промышленных синтезах в качестве побочных продуктов и утилизируются в настоящее время путем сжигания. Поэтому созданная база данных может быть полезной специалистам, занимающимся созданием ресурсо- и энергосберегающих технологий и разработкой рациональных способов переработки отходов. База охватывает литературные источники за период с 1980 по 1999 год и включает в себя около 400 источников.

Система «Лавсан» имеет структуру, аналогичную БД «Lactam». Основное отличие составляют ключевые слова 1-го уровня. Они включают названия толуиловых кислот и их метиловых эфиров: ортолуиловая кислота, метатолуиловая кислота, паратолуиловая кислота, метиловый эфир ортолуиловой кислоты, метиловый эфир метатолуиловой кислоты, метиловый эфир паратолуиловой кислоты. Для создания всей базы «Лавсан» используется 76 ключевых слов-признаков.

Единая внутренняя структура и общие принципы функционирования баз данных «Лавсан» и «Lactam» позволят в дальнейшем легко их объединить, создав общую информационную систему по производству мономеров в химической промышленности Республики Беларусь.

БАЗА ДАННЫХ «КАРБАМИД»

Одним из перспективных направлений развития химической промышленности Республики Беларусь может стать создание технологий синтеза большого числа производных мочевины с использованием карбамида, выпускаемого Гродненским ПО «Азот». Производные карбамида обладают необычайно широким спектром различных потребительских свойств. Фенилалкилкарбамиды применяются в качестве регуляторов роста растений, гербицидов, инсектицидов, бактерицидов, фунгицидов. Алкилпроизводные карбамида применяются для синтеза фармацевтических препаратов, красителей. Диалкилкарбамиды входят в состав смягчителей и антистатиков, применяемых при стирке текстильных изделий. Замещенные карбамиды служат исходными реагентами в синтезе уретанов и полиуретанов, для получения клеевых композиций. Фенилзамещенные карбамиды применяются для изготовления светочувствительной эмульсии. Замещенные карбамиды являются присадками к топливам, маслам, полимерам, повышая их термоокислительную стабильность. Моноалкилкарбамиды с общим числом углеродных атомов от 5 до 25 применяются для удаления нагаров, шлаков и других отложений, образующихся на стенках двигателей внутреннего сгорания. Алкилфенилкарбамиды понижают температуру застывания дизельных топлив. Производные карбамида используются в качестве противостарителей и термостабилизаторов синтетических каучуков, способны предотвращать сшивание полимеров с высокой степенью ненасыщенности. Импрегнирование полиамидных волокон растворами, содержащими карбамид, 1,1-дифенилкарбамид и другие, придает изделиям огнестойкость [15].

Информация о способах производства замещенных мочевины, их применении и потребительских качествах, физико-химических, термодинамических, транспортных и других свойствах веществ, участвующих в процессе производства, химических реакциях с их участием, по методам синтеза и анализа, по технологии, аппаратуре и экономике производства стала основой для создания базы данных «Карбамид».

При формировании базы данных были четко очерчены ее границы, что позволило избежать перегрузки базы большим объемом второстепенных или малозначительных работ по данному вопросу. Так, например, в систему, обслуживающую производство производных карбамида, не включались работы, относящиеся к технологии производства и применения самого карбамида, по-

сколькo объем информации по этим вопросам гораздо больше совокупного объема сведений по технологии производства и применения его производных.

При создании базы данных «Карбамид», также как и в случае БД «Lactam», были определены типы информационных источников и способы работы с ними. Отобрано и систематизировано около 1000 первичных библиографических источников по производным карбамида, составлен список из 150 ключевых слов-признаков, на основании которых сформирована структура базы данных. К основным разделам этого списка относятся: применение замещенных мочеви́н; методы их синтеза; свойства этих веществ; химические реакции с их участием; теория физико-химических процессов; технология и экономика производства; методы анализа производных карбамида; экология и охрана труда. Каждый из этих разделов включает в себя дополнительные подразделы, детализирующие содержание информационного источника.

Раздел «Применение замещенных мочеви́н» содержит 50 слов-признаков, например аграрное применение, аналитическая химия, косметические препараты, лекарственные и физиологически активные препараты, пищевые добавки, поверхностно-активные вещества, присадки к топливам и маслам и т. д.

В созданной базе информация систематизирована на основе разветвленной структуры с использованием ключевых слов различного уровня (см. рис. 7).

Схема базы «Карбамид» состоит из 4 блоков. В блоке I ключевым словом 1-го уровня является «Применение». Ключевые слова 2-го уровня: 1.1. аграрное применение, 1.2. аналитическая химия, 1.3. загустители смазок, 1.4. косметические препараты, 1.5. лекарственные и физиологически активные препараты, 1.6. обработка тканей и волокон, 1.7. отвердители эпоксидных смол, 1.8. пищевые добавки, 1.9. поверхностно-активные вещества, 1.10. полимерные материалы, 1.11. присадки к топливам и маслам, 1.12. растворители, 1.13. стабилизаторы красок и чернил.

Ключевые слова 2-го уровня также дифференцируются.

Блок II представляет собой систему описания вещества различными способами.

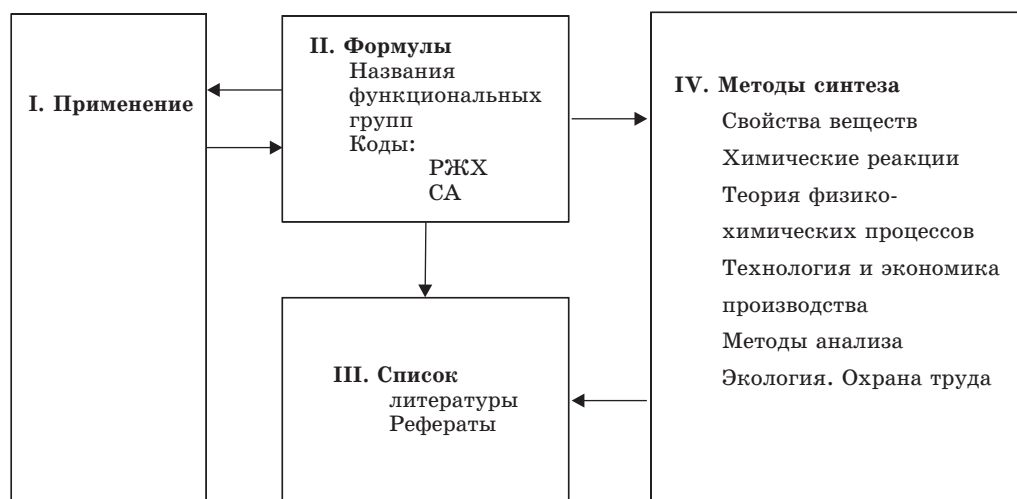


РИС. 7. Схема базы данных «Карбамид»

Регистрация веществ в базе данных проводится по брутто-формулам, названиям, функциональным группам, регистрационным номерам реферативных журналов «Химия» и «Chemical Abstract».

Первую группу регистрации составляет формульная запись. Основой для нее служит эмпирическая формула вещества, составленная по системе Хилла.

Следующий уровень регистрации – названия веществ. Названия веществ приводятся в алфавитном порядке согласно правилам IUPAC. Отдельный список образуют технические, медицинские, фирменные и другие названия соединений.

Особую группу регистрации составляют функциональные производные и продукты замещения. Функциональные производные получаются в результате замены атомов водорода функциональной группой. Их не следует путать с продуктами замещения атомов водорода нефункциональными заместителями: галогенами, нитро-, нитрозо- или азидогруппами.

Последнюю группу регистрации составляют регистрационные номера реферативных журналов «Химия» и «Chemical Abstract». Введение регистрационных номеров журнала «Chemical Abstract» связано с их универсальностью и

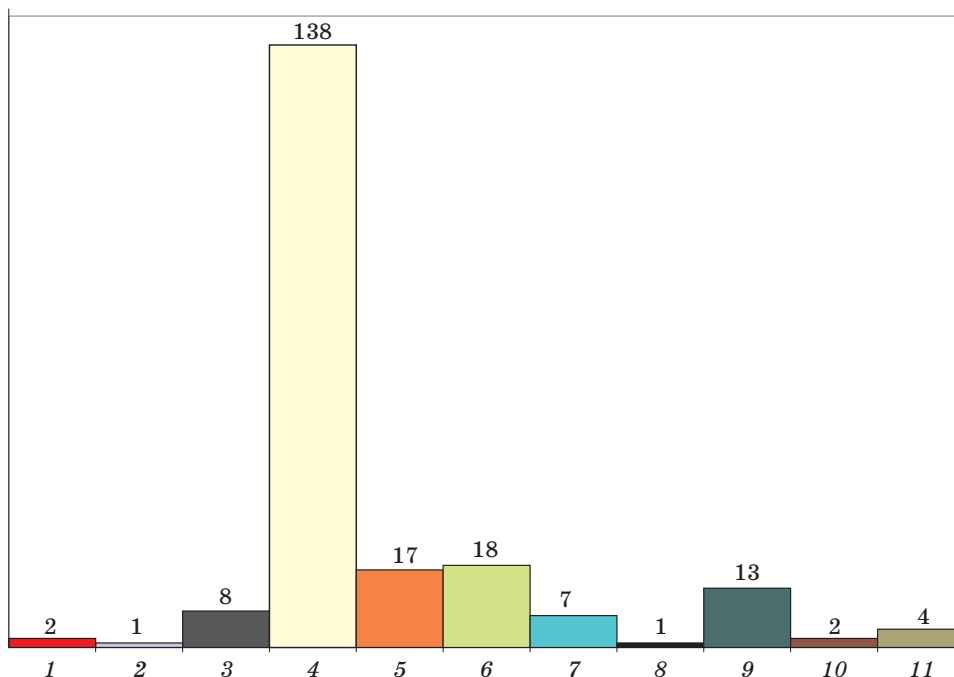


РИС. 8. Распределение числа литературных источников по методам синтеза замещенных мочевины:

1 – амилолиз первичных аминов, 2 – взаимодействие аминов и алкилгалогенидов с цианатами щелочных металлов, 3 – взаимодействие аминов с мочевиной и нитромочевиной, 4 – взаимодействие аминсоединений с органическими изоцианатами, 5 – взаимодействие мочевины или ее производных с различными соединениями, 6 – взаимодействие первичных и вторичных аминов с фосгеном, 7 – взаимодействие с карбамомоилхлоридами, 8 – взаимодействие формамидов с тиокарбаматами, 9 – карбонилирование аминов до замещенных мочевины, 10 – пиролиз карбаматов, 11 – синтез замещенных мочевины из амидов кислот перегруппировкой Гофмана

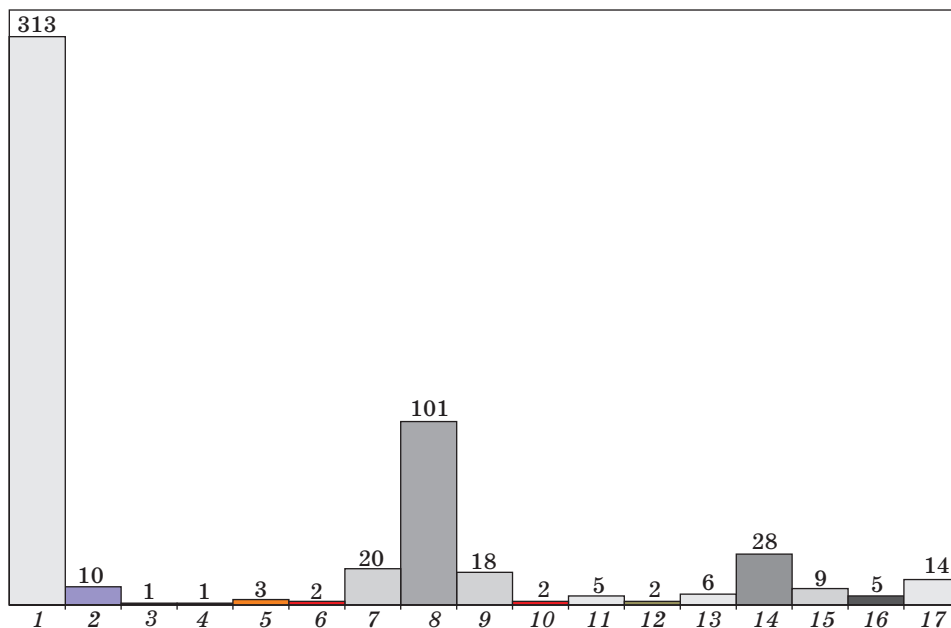


РИС. 9. Распределение числа литературных источников по областям применения замещенных мочевины:

1 – аграрное, 2 – аналитическая химия, 3 – взрывчатые вещества, 4 – водонепроницаемая пропитка бумаги, 5 – дубители, 6 – косметические препараты, 7 – краски, чернила и добавки к ним, 8 – лекарственные и физиологически активные препараты, 9 – обработка тканей и волокон, 10 – огнегасящие и пеногасящие добавки, 11 – отвердители эпоксидных и фенольных смол, 12 – пищевые добавки, 13 – поверхностно-активные вещества, 14 – полимерные материалы, 15 – присадки к топливам и маслам, 16 – растворители, 17 – уретаны

широтой использования в научной, справочной литературе (Sigma, Aldrich). Регистрационные номера располагаются в порядке их возрастания.

Блок III базы данных «Карбамид» представляет набор текстов рефератов публикаций, патентов, статей, отчетов в электронном виде.

Ключевые слова 1-го уровня, составляющие блок IV: «Методы синтеза», «Свойства веществ», «Химические реакции», «Теория физико-химических процессов», «Технология и экономика производства», «Методы анализа». В свою очередь, каждому слову 1-го уровня соответствует набор ключевых слов 2-го и более низких уровней.

Четыре первичных блока, содержащих ключевые слова 1-го порядка, формируют базис информационной системы. Ключевые слова более высоких уровней являются достаточно консервативными, но они могут изменяться по решению группы экспертов, составляющих базу данных.

База данных «Карбамид» на данный момент охватывает литературу по методам синтеза замещенных мочевины и их применению с 1969 по 1990 г. и включает 540 рефератов статей. Распределение их по методам синтеза представлено на рис. 8, а по применению – на рис. 9.

В базе данных «Карбамид» зарегистрировано 235 соединений. Первичными источниками информации в ней в основном являются патенты – 78 %,

журналы составляют 16 %, на другие источники (сборники статей, рукописи и т. д.) приходится 6 % ссылок.

Программное решение для базы данных «Карбамид» представляет собой систему, реализованную в классической, двухуровневой, архитектуре клиент – сервер. В текущей версии системы используется SQL-сервер Firebird 0.9.4 Test1 для Microsoft Windows. Это более современная платформа, чем используемая для информационно-аналитической системы «Lactam».

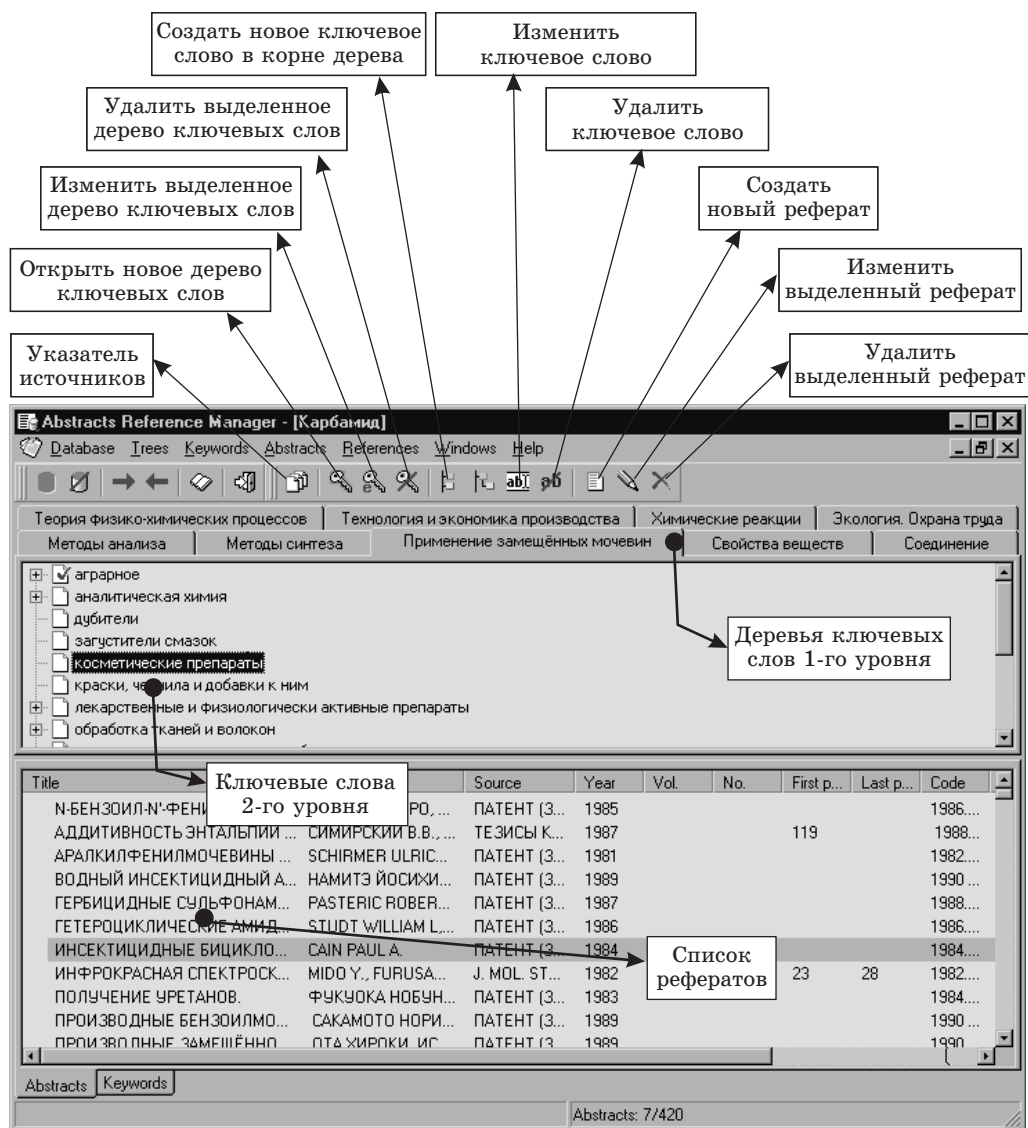


РИС. 10. Поле ключевых слов и рефератов

Концептуально модель базы данных относится к классу объектографических моделей, представляемых в терминах типов объектов, их свойств (аспектов описания) и отношений между ними. Концептуальный уровень описания предметной области является наиболее высоким уровнем абстракции, связывающей неформальные представления пользователей о сущностях реального мира с их информационными образами в создаваемой автоматизированной системе. Эта модель физически реализована в виде взаимосвязанных таблиц. Общий вид окна для ввода и поиска информации представлен на рис. 10.

Для введения больших объемов однотипной информации, например из электронного реферативного журнала ВИНТИ, используется дополнительное программное обеспечение: ArefMgr. В результате текстовая информация из электронного реферативного журнала «Химия» преобразуется в формат HTML, который затем автоматически переносится в базу данных. Кроме этого формат HTML, используемый для предварительной подготовки информации, позволяет легко вводить в базу данных любые источники из Internet.

БАЗА ДАННЫХ «ПЛАСТИЧЕСКИЕ КРИСТАЛЛЫ ОРГАНИЧЕСКИХ ВЕЩЕСТВ»

Предложенное программное обеспечение было использовано при создании базы данных по физико-химическим свойствам пластических кристаллов органических веществ: «Пластические кристаллы органических веществ», являющейся примером разработки узкоспециализированной системы.

Вошедшая в ее основу информация систематизирована в виде разветвленной структуры с использованием ключевых слов различного уровня.

Ключевые слова 1-го уровня включают в себя разделы:

- вещества и системы;
- физико-химические свойства;
- теоретические исследования.

Ключевые слова 2-го уровня разбивают разделы 1-го уровня:

● «Вещества и системы» – органические соединения, ионные соединения, двухкомпонентные системы;

● «Физико-химические свойства» – термодинамические свойства, термофизические свойства, спектры, структура кристаллов, обзоры по физико-химическим свойствам пластических кристаллов;

● «Теоретические исследования» – моделирование пластических кристаллов, интерпретация молекулярной неупорядоченности в пластических кристаллах.

Ключевые слова 3-го уровня расчлняют информацию 2-го уровня на несколько разделов. Например, «Органические соединения»: углеводороды, кислородсодержащие органические соединения, азотсодержащие органические соединения, галогенсодержащие органические соединения, другие; «Термодинамические свойства»: теплоемкость, параметры фазовых переходов, термодинамические функции, давление насыщенного пара и энтальпии сублимации.

Некоторые ключевые слова 3-го порядка дифференцируются ключевыми словами 4-го уровня, например, «Углеводороды»: алканы, алкены, алкины,

циклоалканы, циклоалкены, клеточные углеводороды, ароматические углеводороды.

В пределах ключевых слов 4-го уровня соединения систематизированы по брутто-формуле, например, «Клеточные углеводороды»:

- $C_{10}H_{16}$ адамантан;
- $C_{13}H_{22}$ триметиладамантан;
- $C_{14}H_{20}$ диамантан;
- C_{60} фуллерен;
- C_{70} фуллерен;
- C_8H_{14} бицикло[2.2.2]октан.

БАЗА ДАННЫХ «СТАНДАРТНЫЕ ХИМИЧЕСКИЕ ЭКСЕРГИИ С, Н, О, N, S-СОДЕРЖАЩИХ ОРГАНИЧЕСКИХ ВЕЩЕСТВ»

Отдельного внимания заслуживает рассмотрение базы данных по эксергетическим функциям органических веществ. Эксергетический анализ является перспективным методом оценки степени термодинамического совершенства химико-технологических процессов и аппаратов и служит теоретической базой для разработки энергосберегающих технологий. Имеющиеся базы данных химических эксергий углеводородов и O, N, S-содержащих компонентов нефти ограничены. В настоящее время не существует справочных публикаций по эксергиям широкого класса органических веществ. Таблицы химических эксергий в монографии Шаргута и Петела [16] и учебном пособии Уолла [17] носят скорее демонстрационный характер и содержат небольшое число соединений (порядка 300). Среди получивших распространение изданий следует отметить также монографию Степанова [18], посвященную в основном химическим эксергиям неорганических материалов. Их расширение возможно на основе чрезвычайно дорогостоящих измерений или расчетных процедур, аналогичных применяемым в химической термодинамике.

Созданная нами база данных позволит дать научное обоснование наиболее рациональным способам химической переработки фракций нефти.

Для выполнения эксергетического анализа необходимы численные значения химических эксергий соединений, входящих в состав топлив и химически перерабатываемых фракций нефти. Они были обобщены и систематизированы в специализированной базе данных «Стандартные химические эксергии С, Н, О, N, S-содержащих органических веществ» [19].

Функция эксергии однозначно определяется параметрами окружающей среды, составом рассматриваемого соединения и его термодинамическими свойствами: энтальпией образования, энтропией при температуре окружающей среды и зависимостью теплоемкости от температуры. При фиксированных параметрах окружающей среды для расчета химической эксергии вещества необходимо располагать его термодинамическими характеристиками. Набор соединений, для которых возможны расчеты химической эксергии, определяется наличием для них экспериментальных термодинамических данных, а справочная база данных химических эксергий является надстройкой над базами данных термодинамических свойств.

В данной базе данных использованы термодинамические характеристики свойств веществ: физические свойства соединений, характеристики фазовых переходов, термодинамические свойства веществ в конденсированном состоянии, термодинамические свойства веществ в состоянии идеального газа в интервале температур 0–3000 К [20, 21]. В [20] приведены данные для более чем 3700 углеводородов, а в [21] содержится информация по 2700 соединениям 160 различных классов.

При отсутствии необходимых значений термодинамических величин для расчета эксергии отдельных веществ мы обращались к оригинальным работам по измерению термодинамических свойств. В этом случае использовались сведения, представленные в работах Домальского и др. [22, 23]. Эти публикации являются обзорами экспериментальных работ по измерению термодинамических свойств с экспертной оценкой достоверности представленных в статьях результатов.

В основу базы данных «Стандартные химические эксергии С, Н, О, N, S-содержащих органических веществ» положена классическая структура, используемая при создании термодинамических справочников, которая представляет собой систему разного типа электронных таблиц эксергий веществ в зависимости от температуры:

А. значения стандартных энтальпий сгорания и химических эксергий при 298,15 К следующих классов органических веществ: алканы, алкены, алкины, алкилбензолы, циклоалканы, циклоалкены, спирты, альдегиды, кетоны и эфиры;

В. значения стандартных химических эксергий алканов, алкенов, алкинов, алкилбензолов, циклоалканов, спиртов, альдегидов и кетонов в состоянии идеального газа в интервале 298,15–1000 К;

С. значения стандартных химических эксергий алканов, алкенов, алкилбензолов, циклоалканов, спиртов и кетонов в конденсированном состоянии в интервале 10–380 К.

Количество соединений различных классов в базе данных приведены в таблице.

Таблица

**Количество соединений различных классов в базе данных
«Стандартные химические эксергии С, Н, О, N, S-содержащих органических веществ»**

Класс	Тип таблиц		
	А	В	С
Алканы	219	208	24
Алкены	41	38	15
Алкины	43	28	–
Циклоалканы	50	53	12
Циклоалкены	12	–	–
Алкилбензолы	96	43	11
Прочие углеводороды	24	–	–
Спирты	14	24	7

Окончание таблицы

Класс	Тип таблиц		
	А	В	С
Альдегиды	19	8	–
Кетоны	57	30	6
Простые эфиры	142	–	–
Кислоты	11	–	–
Сложные эфиры	276	–	–
Фенолы	7	–	–
N-органика	51	–	–
NO-органика	7	–	–
S-органика	123	–	–
Всего	1192	432	75

На крупных предприятиях органического синтеза и нефтехимии Республики Беларусь, таких как РУП «Химволокно» (г. Могилев), ГП «Азот» (г. Гродно), Мозырский НПЗ, ПО «Нафтан» (г. Новополоцк), образуется значительное количество различных полупродуктов и побочных веществ, относящихся к различным классам органических соединений. Однако необходимые для технологических расчетов значения физико-химических, термодинамических характеристик этих веществ, как правило, известны только для ограниченного круга соединений. Поэтому целесообразно применение аддитивных методов прогнозирования свойств многочисленных промежуточных и побочных продуктов различных химических производств без дополнительных экспериментов, на основании собранных в базах данных величин для ряда ключевых соединений. Такой подход позволяет осуществлять проверку внутренней согласованности имеющихся данных, проводить массовые расчеты свойств неизученных соединений и их смесей, а также компилировать получаемую информацию в виде приложений в существующие базы данных. Для этого к настоящему времени разработаны оригинальные методики, позволяющие рассчитывать термодинамические свойства многих классов органических соединений в широком интервале температур и различных агрегатных состояниях: алканов, алкенов, алканолов и алканолов [24–26], производных карбамида [27–29], ароматических углеводородов, бензолкарбоновых кислот и их метиловых эфиров [30–32].

Информационно-аналитические системы регулярно пополняются новыми сведениями. Для систем «Lactam» и «Лавсан» проводится работа по переводу информации на обслуживание современной СУБД Firebird. В дальнейшем возможно размещение этих баз данных на общедоступном сервере для обеспечения доступа специалистов к информации через сеть Internet.

Созданные информационные системы, кроме своего основного назначения – поиска литературных источников, могут использоваться для статистического анализа информации, в качестве справочников по физико-химическим, термодинамическим, эксергетическим и другим свойствам, а также для научного обоснования развития действующих производств органического синтеза в республике.

ЛИТЕРАТУРА

1. Юрша И. А. // Хим. пром-сть. 2001. № 4. С. 14.
2. Саркисов П. Д. // Хим. пром-сть. 2000. № 1. С. 20.
3. Кабо Г. Я., Симицкий В. В., Козыро А. А., Юрша И. А. // Ресурсосбережение, энергосбережение и компьютеризация в нефтеперерабатывающей и нефтехимической промышленности: Тр. МНПК РЭЖнефтехим. Новополоцк, 1998. С. 48.
4. Промышленность Республики Беларусь. Статистический сборник. Мн., 2002.
5. Электронный реферативный журнал ВИНТИ: <http://www.viniti.ru>.
6. TRC-home page: <http://www.trc.nist.gov/ThermoML.html>
7. Frenkel M., Chirico R. D., Diky V. V. et al. // J. Chem. and Eng. Data. 2003. Vol. 48, P. 2.
8. Денисов Ю. А., Гринберг Е. Е., Авсеев В. В. // Высокочистые вещества. 1994. № 6. С. 118.
9. Кочуров В. Н. Современные базы данных. Мн., 1998. 160 с.
10. Репин А. А., Осин В. Н., Тузовцев В. В. и др. // Тез. докл. XIV Междунар. конф. по химической термодинамике, 2002. С. 481.
11. Трусов Б. Г. // Там же. С. 483.
12. Цветкова Л. Я. // Там же. С. 484.
13. Крук В. С., Антонова З. А., Кабо Г. Я. и др. // Хим. пром-сть. 1999. № 9. С. 44.
14. Попов А. А. Программирование в среде СУБД FoxPro 2.0. Построение систем обработки данных. М., 1996. 352 с.
15. Вишнякова Т. П., Голубева И. А., Глебова Е. В. // Успехи химии. 1985. Т. 54, № 3. С. 429.
16. Шаргут Я., Петела Р. Эксергия. М., 1968. 279 с.
17. Wall G. Exergetics. Norway, Mölndal. 1998. 149 p.
18. Степанов В. С. Химическая энергия и эксергия веществ. Новосибирск, 1990. 162 с.
19. Говин О. В. Химические эксергии С, Н, О, N, S-содержащих органических веществ: Дис. ... канд. хим. наук. Мн., 1999. 165 с.
20. TRC Tables, Hydrocarbons, Thermodynamics Research Center, Texas A&M University System, Texas, USA, 1994.
21. TRC Tables, Non-Hydrocarbons, Thermodynamics Research Center, Texas A&M University System, Texas, USA, 1994.
22. Domalski E. S., Evans W. H., Hearing D. H. // J. Phys. and Chem. Ref. Data. 1984. Vol. 3. P. 1.
23. Domalski E. S., Evans W. H., Hearing D. H. // J. Phys. and Chem. Ref. Data. 1990. Vol. 19. P. 881.
24. Кабо Г. Я., Козыро А. А., Дiky V. V. // J. Chem. and Eng. Data. 1995. Vol. 40, № 1. P. 160.
25. Говин О. В., Кабо Г. Я. // Ж. физ. химии. 1998. Т. 72, № 11. С. 1964.
26. Кабо Г. Я., Говин О. В., Козыро А. А. // Energy. 1998. Vol. 23, № 5. P. 383.
27. Simirsky V. V., Кабо Г. Я., Frenkel M. L. // J. Chem. Thermodyn. 1987. Vol. 19, № 11. P. 1121.
28. Кабо Г. Я., Козыро А. А., Дiky V. V., Simirsky V. V. // J. Chem. and Eng. Data. 1995. Vol. 40, № 2. P. 371.
29. Дiky V. V., Козыро А. А., Кабо Г. Я. // J. Chem. and Eng. Data. 2002. Vol. 47. P. 239.
30. Maksimuk Yu. V., Кабо Г. Я., Козыро А. А. et al. // J. Chem. and Eng. Data. 1998. Vol. 43, № 3. P. 293.
31. Козыро А. А., Максимук Ю. В., Кабо Г. Я. // Ж. прикл. химии. 2000. Т. 73, № 2. С. 199.
32. Максимук Ю. В., Траханов П. К. // Весці НАН Беларусі. Сер. хім. н. 2000. № 1. С. 48.