
**РАЗВИТИЕ БАНКА ДАННЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ ВИНТИ ПО
ХИМИИ И ХИМИЧЕСКОЙ ТЕХНОЛОГИИ:
ОТ СТРУКТУРНЫХ ДАННЫХ К МАТЕРИАЛЬНОЙ ХИМИИ
И ТЕХНОЛОГИИ.**

**3. РЕАЛИЗАЦИЯ ИНТЕГРАЛЬНОГО ПОДХОДА К АДАПТИВНОЙ СИСТЕМЕ ИН-
ФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ В ОБЛАСТЯХ РАЗРАБОТКИ И ПРИМЕНЕ-
НИЯ ХИМИЧЕСКИХ ВЕЩЕСТВ И МАТЕРИАЛОВ¹**

*Н.Н. Афонина, А.Л. Бирюков, Д.И. Гончаренко, В.М. Зацепин,
В.А. Иванченко, Н.Н. Кочанова, А.К. Мартиросов, О.М. Нефедов,
Р.В. Остапчук, М.А. Федоровская*

Всероссийский институт научной и технической информации РАН, Москва
Научно-инженерный центр «СИСТЕХ», Москва (systech@aha.ru)

Рассмотрены вопросы развития банка данных информационных ресурсов ВИНТИ по химии и химической технологии для адекватной современной поддержки фундаментальных и инновационно-технологических разработок в областях химии и химической технологии.

Проведена разработка и экспериментальная апробация программно-технологических средств веб-доступа к базе структурных данных и базе реакций на актуальных массивах данных. Проведена доработка системы регистрации химических структур и реакций с учетом организации веб-доступа, исследовательских прототипов баз данных регистров химических соединений, реакций и смесей, химико-технологических схем, процессов, аппаратов и аналитических методов. Отработаны различные формы сервиса пользователей: генерация отчетов по произвольным и предопределенным запросам, представление отчетов в виде электронных и печатных публикаций, веб-страниц и локальных реферативных, документальных и документально-фактографических баз данных с развитыми поисковыми возможностями и указателями. Этот сервис практически реализован в областях мембранных технологий и химической и биологической безопасности.

Ключевые слова: химические структуры, химические реакции, химические композиции, химические продукты, химико-технологические схемы, процессы и аппараты, мембранные технологии, системы баз данных, WWW-доступ, информационные продукты.

The issues of development of VINITI information resources databank on chemistry and chemical technology for adequate modern support of fundamental and innovation technique development in the field of chemistry and chemical technology are considered. The software-technological means for Web-access to chemical structure and reaction databases are developed and tested on actual data files. The Registry System of Chemical Substances and Reactions, so as the research prototypes of databases of registers of chemical compounds, reactions and mixtures, chemical-engineering schemes, processes, apparatuses and analytical methods are upgraded.

The various forms of user services are developed: the report generation due to arbitrary and fixed queries, the report representation in forms of electronic and printed publications, in Web forms and reference, full-text and factual local databases having advanced search capabilities and indexes. This service is realised practically in membrane technology and chemical and biological safety regions.

Keywords: chemical structures, chemical reactions, chemical compositions, chemical products, chemical-technological process flow diagrams, processes and apparatuses, membrane technology, databases systems, WWW-access.

В результате предыдущих работ [1-7], программно-технологические средства регистров направленных на создание современной химических информационных объектов (химической интегральной (по информационному наполнению) структур, реакций, композиций, и сервисам) адаптивной (развивающейся) системы материалов, химико-технологических схем, информационного обеспечения (АСИО) работ в процессов и аппаратов, аналитических методов), области химии и химической технологии и которые являются прототипом реляционно-смежных областях, были разработаны объектно-ориентированных информационных

¹ Работа частично поддержана грантом РФФИ 01-07-90097в.

технологий формирования крупных информационных ресурсов по химии и химической технологии, базирующихся на унифицированном представлении, классификации и регистрации химических объектов (от структурных данных до материальной химии и технологии) в полнотекстовых документально-фактографических базах данных. Разработка реализована в среде СУБД MS SQL 7.0/ ОС Windows NT/2000.

Созданные автоматизированные рабочие места (АРМ) ориентированы на внедрение в ВИНТИ и других организациях новой технологии переработки первоисточников с экстракцией, формализацией и индексацией фактографических данных в области химии и наук о материалах, в том числе: библиографического описания, реферата и записи ключевых слов; структур химических соединений, реакций и схем реакций, связанных с определенным документом и/или его рефератом; фактографических данных (термов свойств) о химических веществах, материалах и способах получения, приведенных в обрабатываемых информационных источниках.

Сформированные АРМ позволяют выполнять работы как по загрузке и содержательной переработке информации, так и по проведению информационных исследований в области химии, наук о материалах и ассоциированных областях (в частности, связанных с применением химических веществ и материалов и их воздействием на окружающую среду и здоровье).

В настоящей работе рассмотрены вопросы практического применения результатов разработок [1-7] при создании АСИО в конкретных проблемных областях мембранных технологий и химической и биологической безопасности, а также некоторые общие вопросы развития АСИО в областях разработки и применения химических веществ и материалов.

1. Технология производства информационных продуктов и услуг в области разработки и применения химических веществ и материалов

Информационная технология, ориентированная на проблемный анализ и поддержку принятия решений, и соответствующие программно-технологические средства должны

обеспечивать, как минимум, решение следующих основных задач:

- поиск исходной релевантной информации;
- поиск аналоговой информации в смежных областях (расширение исходной информации);
- обобщение полученной информации;
- уточнение полученной информации;
- анализ информации по заданным аспектам/критериям;
- анализ согласованности информации на каждом этапе.

Необходимое качество решения этих задач может обеспечиваться итерационно, в зависимости от сложности поставленной задачи.

В работах [1-7] рассмотрены вопросы создания интегрированной системы баз данных (БД) как ядра АСИО в области разработки и применения химических веществ и материалов. Здесь же кратко рассмотрим некоторые вопросы современной технологии производства информационных продуктов и услуг в данной предметной области, основные блоки которой характеризуются на рис. 1.

Большинство обеспечивающих программно-технологических подсистем блок-схемы рис. 1 в настоящее время прошли опытную эксплуатацию при создании АСИО в конкретных проблемных областях мембранных технологий и химической и биологической безопасности. Разработанный комплекс ориентирован на обработку всех видов научно-технических публикаций мирового информационного потока, как печатных, так и электронных, и на генерацию/производство различных видов публикаций/отчетов и услуг, в частности: оперативных и ретроспективных проблемных реферативно-документальных сборников и научно-технических журналов в печатном и электронном (в виде CD и интернет-версий с поисковым сервисом), публикаций результатов интернет-мониторинга (на CD и в интернет), сводных коллекций (полнотекстовых баз) по изданным журналам/сборникам с объединенными указателями, локальных проблемноориентированных информационно-поисковых систем (на CD и в интернет-доступе).

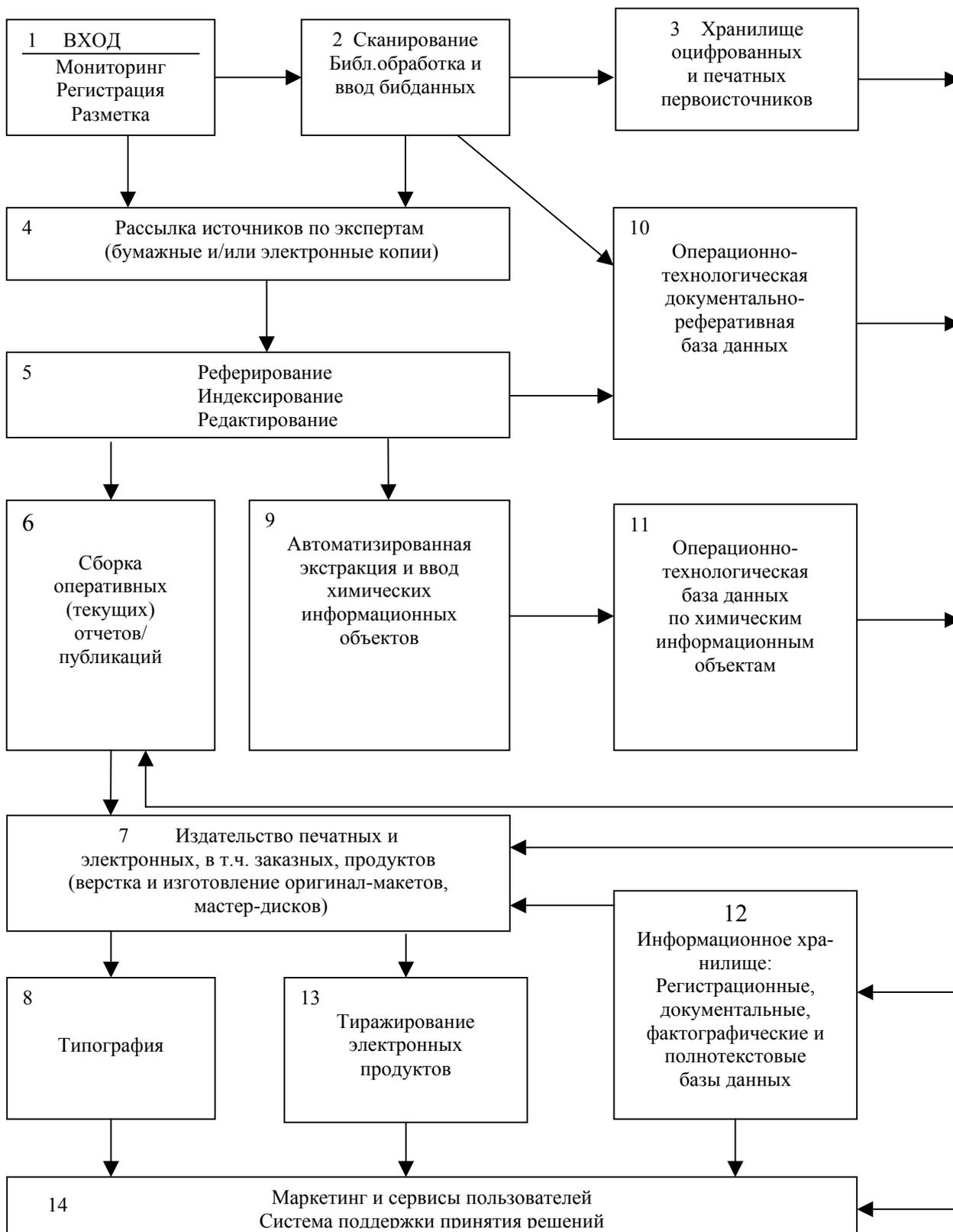


Рис. 1. Блок-схема информационных продуктов и услуг в области разработки и применения химических веществ и материалов

Использование интернет-ресурсов.

Проблемно-ориентированная разработка (добыча информации) интернет-ресурсов, включающая поиск и сканирование веб-

сайтов, их предварительный анализ, классификацию, агрегацию, реорганизацию и т.д., стала в настоящее время важной составляющей в системе информационного обеспечения специалистов в

научно-промышленной сфере. Принципиальными ограничениями в проведении этих работ являются, с одной стороны, большой информационный шум при использовании существующих в настоящее время поисковых машин и, с другой стороны, нестабильность адресов сайтов/URL. Поэтому в любой полноценной системе информационного обеспечения в настоящее время важной остается роль соответствующих вторичных (адаптированных) проблемно-ориентированных ресурсов с расширенными средствами поиска. При этом находимые и используемые первичные интернет-документы/сайты целесообразно фиксировать в виде специально организованных временных срезов – проблемных интернет-ресурсов (ПИР) на компакт-дисках (CD), которые в настоящее время являются наиболее дешевым и эффективным носителем благодаря прогрессу в скоростях доступа.

В рамках настоящей работы было создано необходимое программно-технологическое обеспечение и форма публикации ПИР на CD в виде, максимально приближенном к сеансу работы в Интернет, с возможностью прямого использования URL-ссылок ПИР для перехода в среду Интернет.

Организация веб-доступа к химико-структурным данным. При решении задачи обеспечения веб-доступа к химико-структурным данным (индивидуальным соединениям и реакциям) учитывались следующие основные требования/ возможности: запрос по интересующей пользователя химической структуре; ввод структуры как через Java апплет, так и путем указания файла, где находится структура; сохранение в базе данных запросов пользователей и их результатов; парольная система, идентификация и регистрация пользователей с необходимыми проверками (на заполнение обязательных полей, на уникальность имен пользователей и т.д.) с занесением информации в базу данных; реакция на ошибочные действия пользователя; многопользовательская обработка данных.

Разработанные программно-технологические средства включают компоненты:

1. Внешний СОМ сервер структурной базы данных (предоставляет возможность клиенту послать на запрос структуру молекулы и принять результаты запроса).

2. CGI клиент, который принимает запросы от пользователя, обрабатывает их, отсылает их СОМ серверу и принимает результат запроса, обрабатывает полученный результат и отправляет обратно пользователю в виде HTML страницы. В обработку запроса пользователя входит:

- Определение типа структуры (рисунок, mol-файл – формат MDL).
- Выделение идентификатора пользователя.
- Выделение текстового описания структуры.
- Выделение параметров запроса.

3. Интерфейс пользователя, обеспечивающий регистрацию пользователя, получение пароля по электронной почте, переход на страницу запроса, работу с парольной системой и настройками, в частности выбор опции работы с апплетом для рисования химических структур (апплет будет загружен) или работы с mol-файлом без загрузки апплета. Окно апплета показано на рис. 2.

Апплет разрабатывался с учетом обеспечения

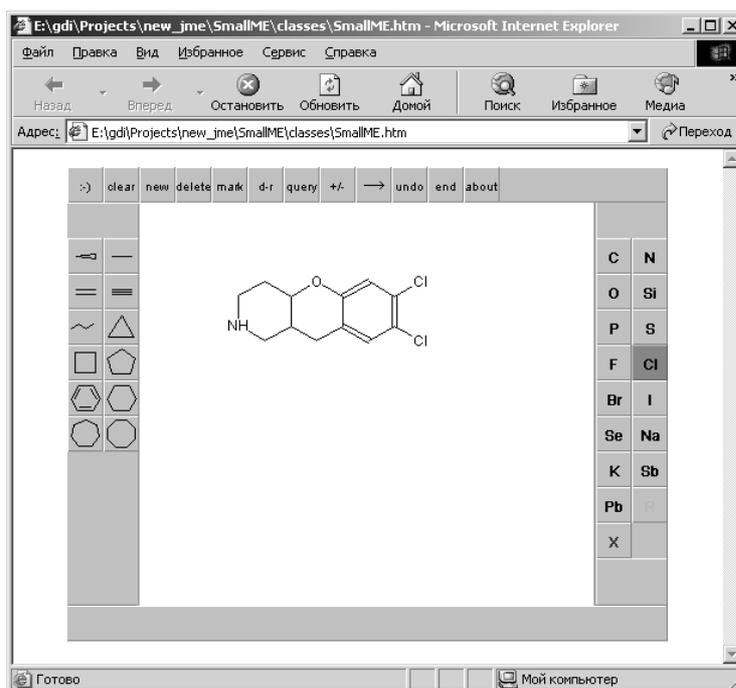


Рис. 2. Окно апплета.

функциональности, реализованной в известных аналогах (Марвин, Новартис, КэмбриджСофт, MDL).

После того, как CGI клиент обработал запрос пользователя, он посылает данные на запрос. После получения результатов запроса, клиент обрабатывает этот результат следующим образом: в зависимости от результатов запроса, клиент выдает пользователю только те данные, которые соответствуют выбранным параметрам. При этом весь присланный результат запроса преобразуется в HTML формат путем добавления необходимых тегов.

2. Информационное обеспечение работ в области мембранных технологий

Мембранные технологии стали в настоящее время важной составной частью химической технологии в целом и являются основой для развития многих технологических областей или направлений исследований и разработок. Достаточно очевидно, что дальнейшее развитие мембранных технологий должно основываться на серьезной информационной базе, учитывающей тенденции развития мирового рынка мембранных технологий.

Своевременное и полное удовлетворение информационных потребностей пользователей предполагает переход к адаптивной (развивающейся) системе информационного обслуживания по мембранным технологиям (АСИО - МЕМБРАНЫ), настраиваемой в соответствии с изменениями внешней среды (новых направлений исследований, новых проблем и методов решения).

Характеристика предметной области.

В мировой практике мембранные процессы нашли наиболее широкое применение в следующих областях:

обработка и очистка воды: обессоливание, получение сверхчистой воды для электронной промышленности, медицины и здравоохранения, водоподготовка в энергетике, бытовые водоочистители, очистка бытовых стоков;

газоразделение: выделение углекислого газа при добыче природного и нефтяного газов, обогащение кислородом воздуха для ме-

дицинских и производственных целей, создание нейтральных сред за счет обогащенного азотом воздуха, извлечение водорода из технологических газовых смесей, создание регулируемой газовой среды для хранения сельскохозяйственной продукции;

химическая промышленность: производство хлора и каустика, концентрирование химикатов, регенерация и очистка растворителей, моющих растворов и масел, селективное выделение ионов металлов и т.п.;

пищевая промышленность: производство молочных продуктов, вина, фруктовых и овощных соков, пива, сахара, кофе, извлечение белков и лактозы из молочной сыворотки, стерилизация, концентрирование, осветление, обессоливание и т.п.;

биотехнология и медицина: стерилизация растворов медицинских препаратов и физиологических растворов, сбор клеточных культур, генная инженерия, очистка и концентрирование биологически активных веществ и лекарственных препаратов, гемодиализ, оксигенация крови и др.;

очистка стоков промышленных предприятий в химии, нефтехимии, машиностроения и др. отраслях народного хозяйства с возвратом в производство ценных продуктов;

газоразделение в процессах нефте- и газопереработки: выделение гелия, водорода, аммиака, углеводородов, оксидов углерода и других газов из газовых смесей;

обессоливание воды (обратный осмос, нанофильтрация, электродиализ, комбинированные схемы) для промышленных предприятий, энергетических объектов и питьевого водоснабжения;

разделение веществ методом испарения через мембрану - первапорация взамен используемых в настоящее время энергоемких, неэкологичных и высокотемпературных процессов ректификации, дистилляции и др.;

предотвращение техногенного нефтяного загрязнения почв и гидросферы и испарения нефтепродуктов и химических веществ, методы определения, разделения и обезвреживания радионуклидов;

селективное выделение индивидуальных компонентов из техногенных и природных вод нетрадиционными методами, в том числе для пополнения сырьевых источников драгоценных, редких и цветных металлов за счет экономичного и эко-

гичного использования высокоминерализованных подземных и морских вод.

Понятийно-терминологическая база и классификация.

Понятийная и терминологическая база предметной области находится в настоящее время в состоянии развития. В общем случае достижения в области производства и применения мембран классифицируются развитием мембранных технологий по сепарируемым средам (газ, жидкость) и по физическим условиям фильтрации (размер пор и давление). В то же время классификация тех или иных секторов экономики в непосредственных исследованиях перспектив рыночных продаж преимущественно определяется развитием мембранных технологий по материалам, используемым для производства активных элементов, форме и конфигурациям мембранных модулей и по промышленным областям применения.

Основы классификации закреплены в документе “Международная терминология по мембранам и мембранным процессам. Рекомендации Международного объединения специалистов по теоретической и прикладной химии IUPAC 1996” [8]. Этот документ разработан ведущими специалистами в области мембранной науки и технологии для обеспечения единообразия научной классификации и улучшения взаимопонимания деятелей науки и техники и исключения ошибок, связанных с неадекватным употреблением терминов. Согласно этому документу, сепарация газовых сред подразделяется на:

1. собственно разделение газов
2. паровое разделение
3. разделение перфузией (первопарация)

Сепарация жидких сред подразделяется на:

1. диализ (электродиализ)
2. нанофильтрацию
3. ультрафильтрацию
4. микрофильтрацию
5. обратный осмос

Источники информации.

Публикации по мембранам (более 30 тыс. в год) распределяются по более, чем 1000 периодических изданий, из которых лишь не-

сколько являются специализированными мембранными журналами, что связано, скорее всего, с использованием и применением мембранных технологий во многих отраслях науки и техники. Патенты являются важной составляющей информационного потока и их доля только по технологическим мембранам составляет около 10–15% от мембранных публикаций. В последние годы отчетливо возрастает и роль интернет-ресурсов, накапливаемых на веб-сайтах исследовательских центров, ряда международных, государственных, общественных организаций и фирм, представляющих широкий спектр научно-технической и коммерческой информации в области мембранных технологий.

Общая архитектура развиваемой АСИО представляет собой симбиоз информационной системы (ИС) генерации ресурса, включающей операционно-технологические базы данных и ИС пользователей (корпоративная или персональная информационная система). ИС генерации ресурса обеспечивает формирование фонда первичной информации (импорт электронных документов и селектов из различных БД, сканирование веб-сайтов, оцифровку бумажных первоисточников и др.), ее переработку (индексацию, реорганизацию, унификацию, интеграцию, накопление), хранение в течение технологического цикла, проведение информационно-аналитических исследований и формирование проблемно-ориентированных информационных выборок. ИС пользователей включает генерируемые проблемно-ориентированные БД и информационные издания в электронных (веб, на компакт-дисках (КД)) и бумажной формах. При этом соответствующие программные оболочки генерируемых электронных изданий и БД обеспечивают иерархически организованные рубрикаторы информации, указатели (словарный, авторский, предметный, источников, формульный) и разнообразные функции пользовательских запросов, в том числе отображение найденной информации, генерацию отчетов по результатам поиска и их выгрузку в файлы или на печать.

В плане практического внедрения АСИО - МЕМБРАНЫ в настоящее время имеется пятилетний опыт подготовки и издания информационно-аналитического журнала «МЕМБРАНЫ» (бумажная, электронная CD- и интернет-версии (<http://www.chem.msu.su/rus/journals/membranes/welcome.html>), периодичность – 4 номера в год).

Создана и постоянно обновляется CD-база данных «МЕМБРАНЫ» по мировому потоку публикаций с 1998г. Подготовлен CD-выпуск «Интернет-ресурсы по мембранам и мембранным технологиям».

Для работы с фактографической и документально-фактографической информацией по мембранным технологиям использованы разработанные программно-технологические решения, описанные в работе [2]. Проиллюстрируем их применение для формализации химико-технологической информации из работы [9] (используемые в работе определения технологической операции (стадии), процесса, аппарата соотв. **Технология (Т)**) до подготовки [9] включает технологические операции (стадии):

1. дозирование в исходную воду специальных ингибиторов осадкообразования в количестве 1–5 мг/л;
2. обработку воды на мембранных установках;
3. регулярные гидравлические промывки

мембранных фильтров со сбросом давления;

4. регулярные химические регенерации мембранных фильтров с помощью специальных растворов;
5. на выходе очищенной воды из установки предусматриваются ультрафиолетовые бактерицидные лампы;
6. Доочистку фильтрами умягчения.

Оригинальное изображение технологической схемы (рис.6 [9]) из электронной версии журнала «МЕМБРАНЫ», представлено на рис. 3.

В операционно-технологической базе данных (ОТБД) по химическим информационным объектам (ХИО) этот чертеж также хранится в виде графического объекта (рис. 4).

В редакторе технологических схем [2] эта технологическая схема представляется в виде функциональной схемы - взвешенного графа (рис. 5), хранимого в базе данных как математический объект. Каждая из вершин этого графа может быть детализована (развернута) в виде текстовых и численных фактов: комплекса аппаратов (операции 1, 2, 5, 6) со своими параметрами и/или

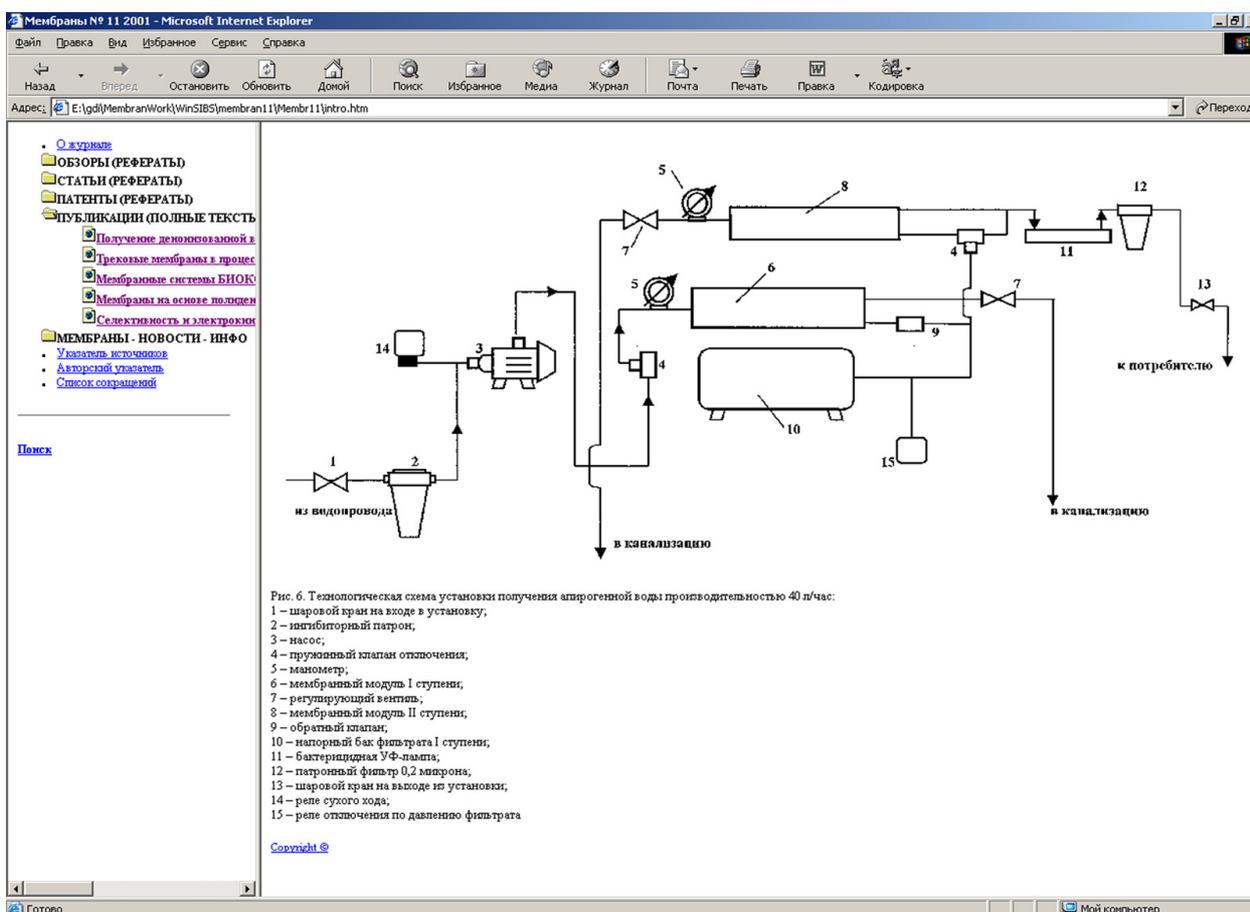


Рис. 3. Электронная версия журнала «Мембраны» (№ 11 за 2001 г.)

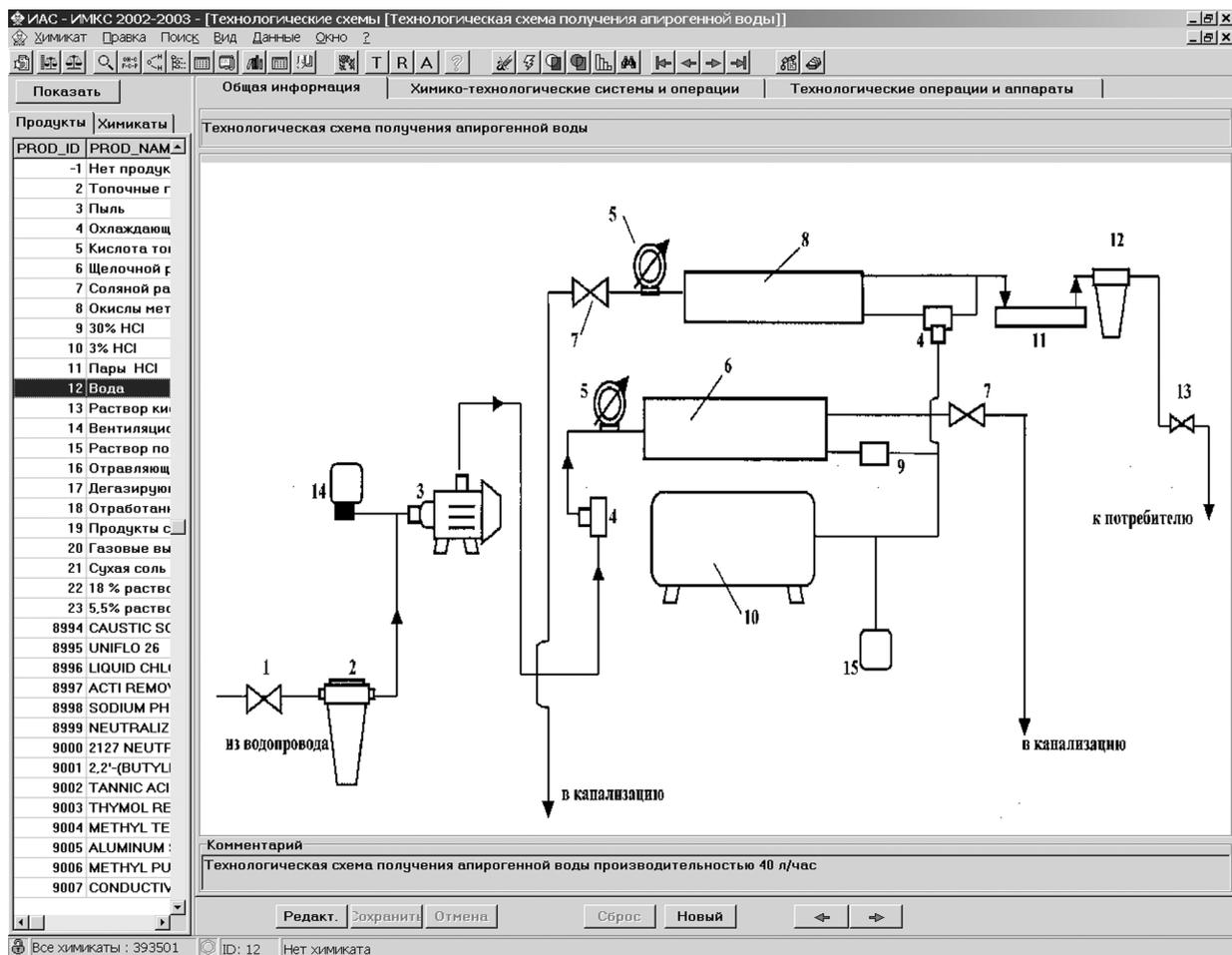


Рис. 4. Интерфейс ОТБД по ХИО, закладка «Общая информация» формы «Технологические схемы» (чертеж технологической схемы)

других релевантных характеристик (например, подробное описание операций 3 и 4). Представление комплекса аппаратов (структурная схема) технологической схемы приведено на рис. 6.

Организация информации в виде системы баз данных, реализованная в среде СУБД MS SQL Server 7.0, обеспечивает полнофункциональный поиск по любым элементам данных и их комбинациям, как фактографических данных, так и полнотекстовой информации.

3. Информационное обеспечение проблем химической и биологической безопасности

На рубеже второго-третьего тысячелетий перед человечеством встала поистине жизненно важная глобальная проблема химической и биологической безопасности. Актуальность ее вызвана не столько негативным влиянием естественных природных явлений, сколько антропогенной деятельностью, став-

шей источником губительного воздействия на биосферу. Поступление в окружающую среду многочисленных вредных химических и биологических факторов обусловлено несовершенством технологических процессов производства промышленной и сельскохозяйственной продукции, электроэнергии и тепла, отсутствием или недостаточной эффективностью очистных сооружений, возможностью аварийного выброса в окружающую среду токсических реагентов при их транспортировке, хранении, переработке и т. д.

Потенциальным источникам химической опасности являются также накопленные в огромном количестве отравляющие вещества, в том числе и сохранившееся химическое оружие, места его хранения и уничтожения, оставленное химическое оружие – могильники, которые располагаются как на суше, так и на дне морей и океанов.

Развитие науки, в том числе биотехнологии, хотя и привело к победе человечества над опасными инфекционными болезнями, создало прецеденты контаминации окружающей среды высоко-

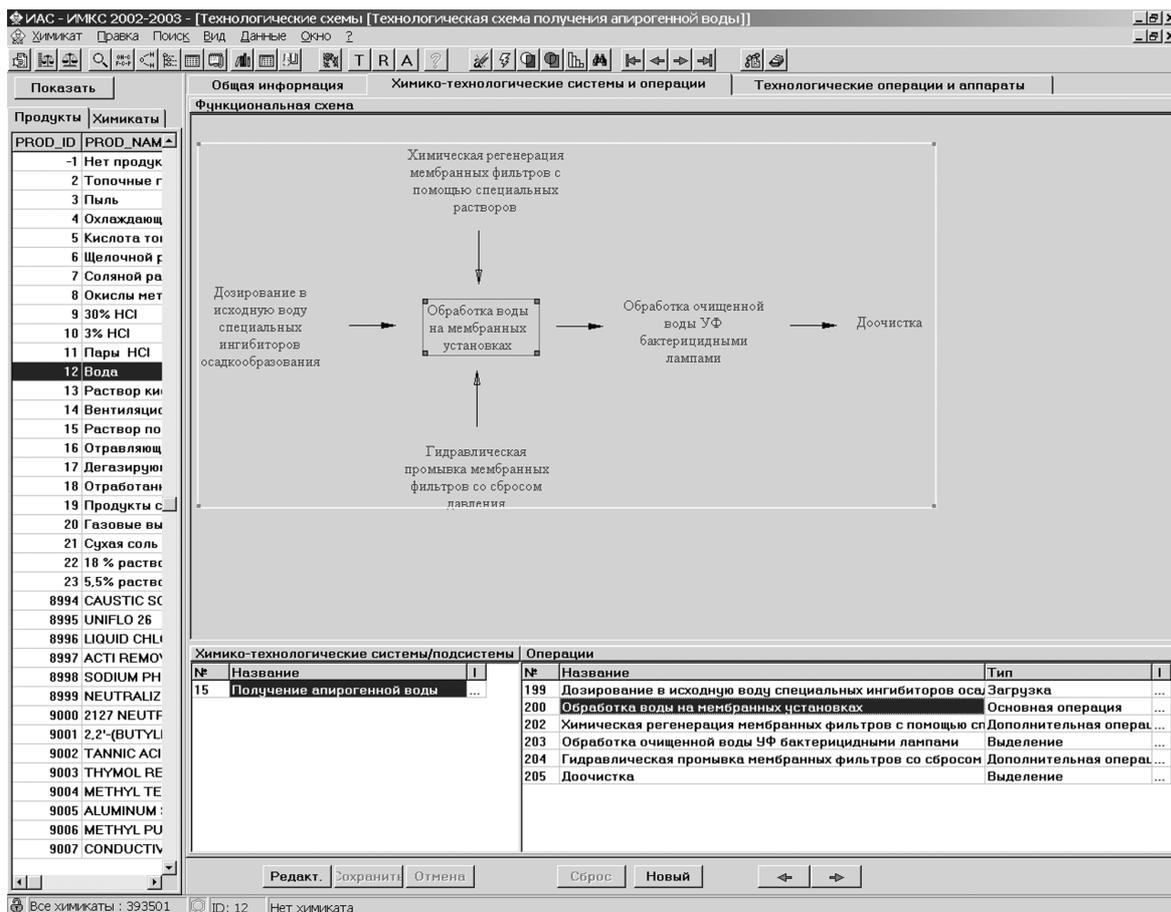


Рис. 5 Граф (функциональная схема) технологической схемы рис.3 и 4, закладка «Химико-технологические схемы и операции» формы «Технологические схемы» (функциональная схема)

вирулентными возбудителями бактериальной, вирусной и риккетсиозной природы. Несовершенство систем биологического контроля при массивном внедрении биотехнологии в производство лечебных и диагностических иммунобиологических препаратов, продуктов питания, биологически активных добавок и т. д. также представляет реальную биологическую опасность для людей.

Несмотря на ратификацию Конвенций о запрещении биологического и химического оружия (1972 и 1994 гг., соответственно), во многих странах продолжают проводиться работы, прямо или косвенно связанные с созданием суперэкоксикантов химической и биологической природы. Существующие запасы отравляющих веществ, наличие источников получения штаммов возбудителей инфекционных болезней, относительная простота и доступность технологии наработки токсических химикатов, прекурсоров и культивирования патогенных штаммов микроорга-

низмов создают реальную опасность использования их в антигуманных целях.

Эффективность и комплексность решения всех аспектов проблемы химической и биологической безопасности во многом определяется доступностью сведений и информированностью специалистов по всем интересующим вопросам.

Развиваемая с использованием общих подходов [1-7] адаптивная система информационного обслуживания по проблемам химической и биологической безопасности (АСИО-ХББ) обеспечивает в настоящее время формирование фонда первичной информации (импорт электронных документов и селектов из различных баз данных, сканирование веб-сайтов, оцифровку бумажных первоисточников и др.), ее переработку (индексацию, реорганизацию, унификацию, интеграцию, накопление), хранение в течение технологического цикла и формирование как различных информационных продуктов, так и предусматривает возможность проведения информационно-аналитических исследований и подготовки других проблемно на-

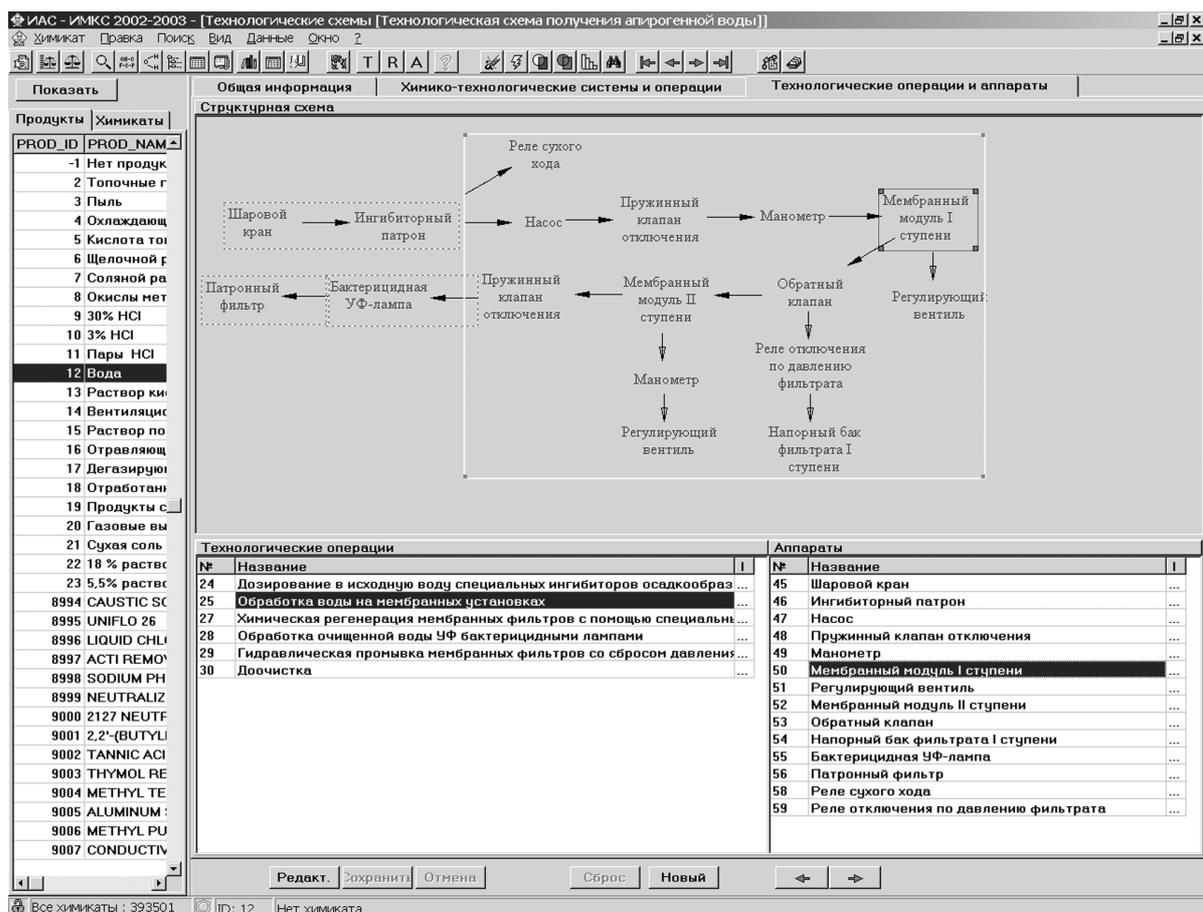


Рис. 6. Интерфейс ОТБД по ХИО, закладка «Технологические операции и аппараты» формы «Технологические схемы» (структурная схема)

страиваемых информационных выборок, продуктов и изданий.

Характеристика предметной области.

Сформировавшиеся за последние годы потоки научных публикаций, патентов и нормативно-правовых документов по этой проблеме рассредоточены по многочисленным информационным медицинским, химическим, технологическим рубрикам по следующим аспектам:

- размещение объектов гражданского, промышленного и сельскохозяйственного назначения и установление их санитарно-защитных зон;
- продукция производственно-технического назначения, товаров для личных и бытовых нужд и технологии их производства;
- работы с потенциально опасными для человека химическими веществами, патогенными микроорганизмами, токсическими продуктами их жизнедеятельности и отдельными видами продукции;

- производство пищевых продуктов, биологически активных добавок, продовольственного сырья, контактирующих с ними материалов и изделий;
- импорт на территорию Российской Федерации продуктов питания, фармацевтической промышленности, бытовой химии и т. д.;
- эксплуатация водных объектов, обеспечение промышленного и бытового водоснабжения;
- контроль загрязнения атмосферного воздуха в городах и сельских районах, на территориях промышленных предприятий, в рабочих зонах производственных помещений.

Существенное место в этих информационных потоках занимают:

- научно-технические и нормативно-правовые аспекты санитарно-эпидемиологического и гигиенического нормирования или категорирования химической и биологической опасности;
- выявление, изучение и классификация источников химической и биологической опасности;

- выявление и обоснование для лицензирования отдельных видов деятельности (работ, услуг), представляющих потенциальную химическую и биологическую опасность;

- доклинический отбор химиотерапевтических и иммунобиологических средств и препаратов, их фармакология и фармакокинетика, нежелательные последствия при использовании лекарств;

- токсикологические характеристики элементоорганических и органических соединений, промышленных загрязнителей, профессиональные отравления;

- научно-производственные аспекты химической и биологической технологии получения физиологически активных веществ.

Подсистема формирования информационных ресурсов.

Общая схема формирования информационных ресурсов, осуществляемого в рамках АСИО-ХББ, приведена на рис. 7.

Информационный мониторинг по проблемам химической безопасности охватывает как достаточно обширные потоки первичных научных публикаций, патентов и нормативно-правовых документов, так и вторичных информационных ресурсов, рассредоточенных по многочисленным медицинским, токсико-

логическим, экологическим, химическим, технологическим рубрикам информационных изданий и баз данных. Классификация соответствующего информационного ресурса, выявленного и накапливаемого в ходе начальных этапов практической реализации АСИО-ХББ, включает следующие рубрики:

1. Общие вопросы химической безопасности.

- 1.1. Нормативные акты (конвенции, протоколы, соглашения), конференции, симпозиумы по проблеме химической безопасности.

- 1.2. Проблемы реализации Конвенции о запрещении химического оружия.

- 1.3. Общие методологические аспекты, моделирование и оценка риска химической опасности.

- 1.4. Санитарно-гигиенические нормы и специальная техника безопасности.

- 1.5. Нормативно-правовая справочная информация.

2. Источники химической опасности.

- 2.1. Токсические химикаты, ядовитые и аварийно опасные химические вещества.

- 2.2. Объекты хранения и уничтожения химического оружия и другие опасные химические предприятия и производства.

- 2.3. Нежелательные эффекты лекарственных и фармакологических препаратов, пестицидов и средств бытовой химии.

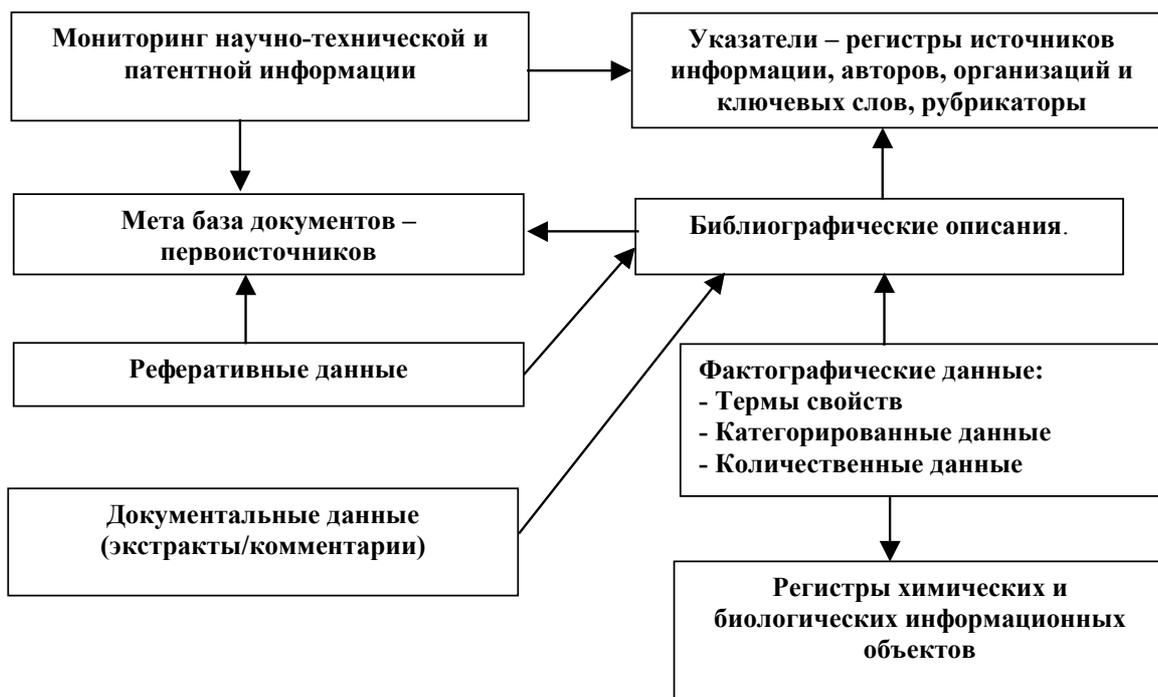


Рис. 7. Общая схема формирования информационных ресурсов в виде интегральной многоаспектной базы данных, осуществляемого в рамках проекта АСИО-ХББ.

2.4. Множественная химическая чувствительность, биологическое действие сверхмалых доз химических веществ.

3. Химический терроризм.

3.1. Общие вопросы предотвращения использования химических веществ в террористических целях.

3.2. Выявление и контроль возможных путей получения отравляющих веществ.

3.3. Прецеденты отравлений, химические аварии и других чрезвычайные ситуации, связываемые с вероятными актами проявления химического терроризма.

3.4. Медико-биологические и социально-психологические последствия химических террористических акций.

4. Средства и методы выявления химической обстановки и индикации сильнодействующих, ядовитых и отравляющих веществ.

4.1. Общие вопросы осуществления мониторингового контроля химической опасности.

4.2. Физико-химические, биокаталитические, иммунологические и токсикологические методы выявления и индикации токсических веществ.

4.3. Аналитические и информационные системы, комплексы, приборы, сенсорные и биосенсорные устройства, пригодные для мониторинга и выявления факторов химической опасности.

5. Методы и средства технической защиты, профилактики и лечения.

5.1. Общие вопросы технической и медицинской защиты от поражающих факторов химической природы.

5.2. Средства защиты органов дыхания, глаз, кожи.

5.3. Средства, способы и методы деконтаминации объектов окружающей среды.

5.4. Средства, способы и методы профилактики, лечения и антидотной терапии.

6. Ликвидация последствий химического заражения.

6.1. Общие вопросы, связанные с ликвидацией последствий химического заражения.

6.2. Расследование прецедентов химического заражения.

6.3. Средства, способы и методы деконтаминации объектов окружающей среды.

Метабаза документов – первоисточников, постоянно пополняемая в ходе мониторинга научно-технической, патентной информации и баз данных, содержит описания (резюме/рефераты) источников данных и служит основанием для поиска/заказа первоисточников и формирования соответствующих записей URL/Библиографического описания в интегральной многоаспектной базе данных ХББ.

Библиографическое описание может представлять собой первичную ссылку на регистрируемый документ и/или вторичную ссылку на его цитируемый фрагмент/комментарий.

Реферативные данные, получаемые в ходе содержательной переработки первоисточников непосредственно связаны с соответствующими записями не только библиографических описаний, но и мета базой документов.

«Указатели ...» - система постоянно обновляемых и пополняемых справочников облегчает как содержательную обработку документов (регистрацию источников информации, авторов, организаций, формирование записей ключевых слов и рубрик документов), так и организацию доступа к информации в АСИО ХББ.

Документальные данные (экстракты/комментарии) включают записи отдельных фрагментов документов с библиографическими описаниями.

Фактографические данные (термы свойств, категоризованные данные и количественные данные) формируются по результатам содержательной переработки документов.

Регистры химических и биологических информационных объектов представляют собой особый тип формализуемых и регистрируемых объектов (химических структур, продуктов, реакций, процессов, аппаратов и других объектов живой и неживой природы), которым в ходе содержательной переработки информации приписываются определенные фактографические данные.

Основные операции содержательной переработки информации.

Содержательная переработка информации в рамках АСИО-ХББ включает следующие основные операции:

- формализация структурированных документов по элементам данных (библиография, резюме/реферат, содержательные разделы тек-

ста, ключевые слова, графики, фактографические данные и др.) и их регистрация в БД;

- экстракция и/или ввод и формализация элементов данных обрабатываемых документов;
- реферирование (подготовка и редактирование реферата и ключевых слов поискового образа документов (ПОД));
- экстракция, формализация, индексация и/или набивка химико-структурных данных (химических структур, в том числе индивидуальных и формул Маркуша, реакций и термов их свойств);
- формализация и унификация фактографических данных.

Формализация и регистрация информационных объектов из сферы материальной химии служат основанием для индексации релевантных документальных, графических и фактографических данных. Фактографические данные на начальном этапе формализуются терминами свойств, которые на последующих этапах унифицируются в релевантные качественные/категорированные и количественные данные.

Специфика подготовки реферативных изданий/разделов.

Содержательная переработка информации для выпуска реферативных изданий и реферативных разделов комбинированных продуктов включает операции:

- сканирование и распознавание текстовых фрагментов первоисточников; формирование исходного массива (файлов) библиографических данных с присоединенными резюме/рефератами/текстовыми фрагментами и/или первоисточниками (при необходимости и на бумажных носителях);
- подготовка и включение в электронный массив записей: перевода основного названия, реферата и ПОД;
- проведение поиска в отечественных (ВИНИТИ) и зарубежных базах данных (Medline, SCI и др.) и формирование дополнительного массива "подходящих" рефератов (обычно 20-30% от общего количества);
- загрузка объединенного рабочего массива в операционно-технологическую БД со структурированием записей по элементам

данных (в настоящее время – в соответствии с действующим в ВИНИТИ НТП);

- проведение дополнительного редактирования записей перевода основного названия, реферата и ПОД;
- подготовка и макетирование нереферативных разделов изданий (материалы от редакции или отдельных авторов по представлению членов редакционного совета);
- генерация оригинал-макета реферативного раздела, указателя источников, авторского указателя и пермутационного предметного указателя по запросу из БД;
- формирование оригинал-макета текущего выпуска издания, распечатка и контрольное редактирование, распечатка контрольных экземпляров выпуска для согласования/утверждения;
- внесение изменений, подготовка окончательного варианта оригинал-макета бумажной версии издания, тиражирование выпуска в типографии;
- подготовка электронной версии издания и тиражирование ее на CD;
- получение тиража издания из типографии, рассылка/передача авторских экземпляров членам Редакционного совета и Редколлегии;
- подготовка и рассылка рекламных материалов, участие в оформлении подписки, рассылка/передача издания подписчикам.

Формируемый в ходе информационного мониторинга и содержательной переработки первичных и вторичных источников информации ресурс оформлен в виде операционно-технологической БД (ОТБД) для формирования вторичных информационных продуктов в электронной и печатной формах. Пример экранной формы пользовательского интерфейса (ОТБД) для реферирования документов приведен на рис. 8.

На рис. 9 приведен пример экранной формы пользовательского интерфейса ОТБД для генерации оригинал-макета реферативного раздела, указателя источников, авторского указателя и пермутационного предметного указателя по запросу из БД.

Информационные продукты и услуги АСИО-ХББ.

В настоящее время выпускаются:

- оперативные и ретроспективные проблемные реферативно-документальные сборники;

• информационно-аналитический журнал «Химическая и биологическая безопасность», первоначально выпускавшийся в виде реферативного журнала, преобразован в комбинированный журнал со структурой: обзоры, оригинальные публикации+ рефера-

тивный раздел в печатном и электронном (в виде CD и интернет-версий (www.cbsafety.ru) с поисковым сервисом);

• публикуются результаты интернет-мониторинга (ПИР на CD).

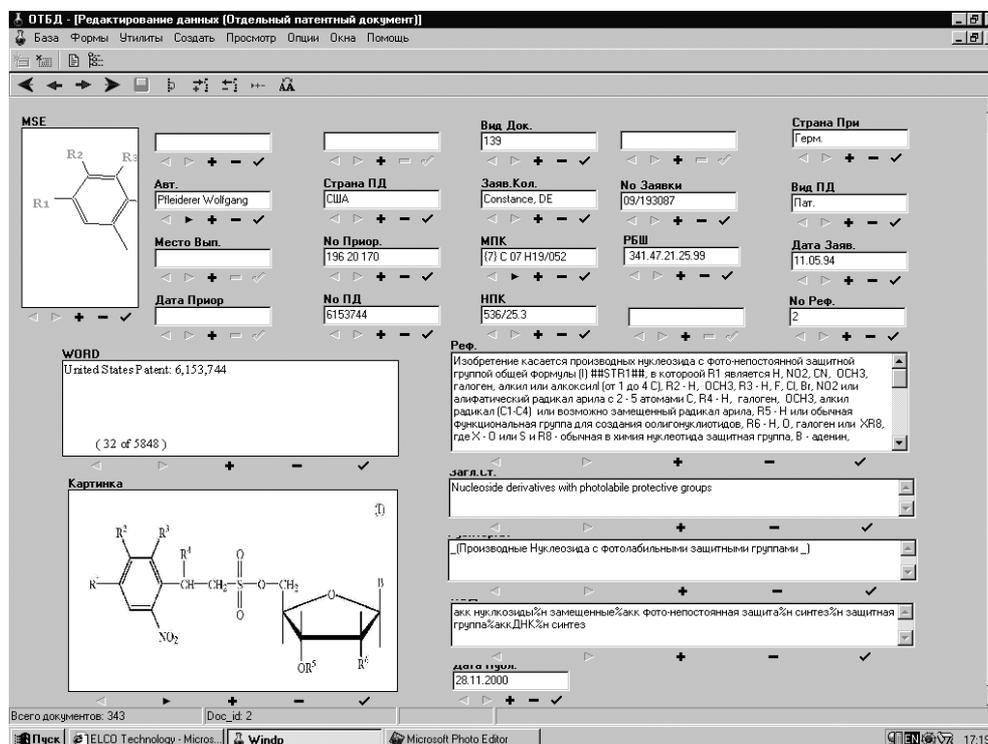


Рис. 8. Экранная форма пользовательского интерфейса ОТБД для реферирования документов (патентов).

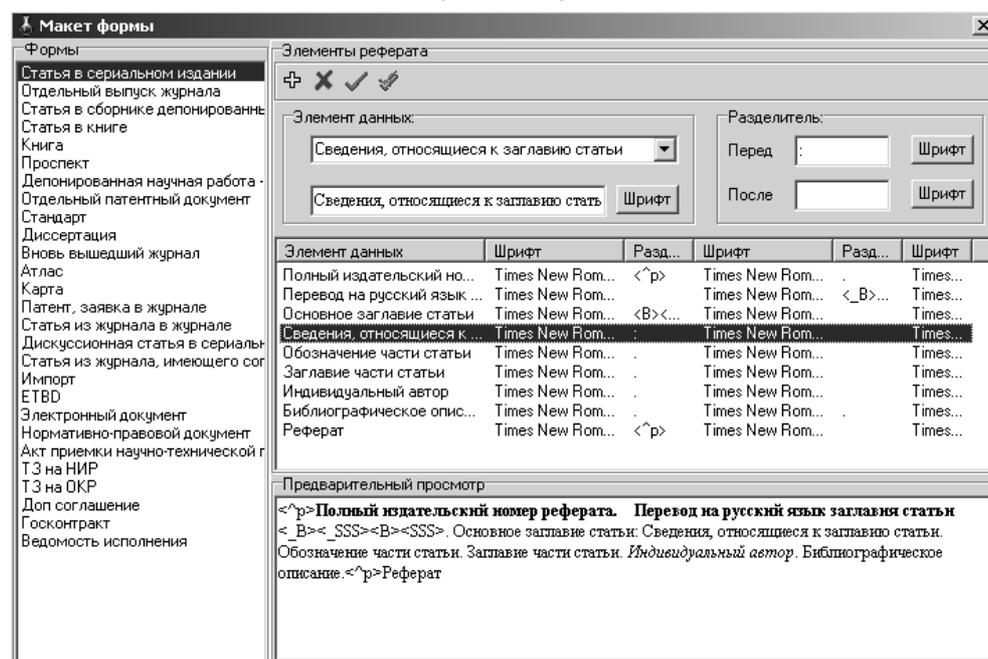


Рис. 8. Экранная форма пользовательского интерфейса ОТБД для генерации оригинал-макета реферативного раздела, указателя источников, авторского указателя и пермутационного предметного указателя по запросу из БД

Основную часть АСИО-ХББ составляет система фактографических БД по химическим и биологическим информационным объектам, разработанная в соответствии с подходом работ [1-7, 11-13] и являющаяся основой системы поддержки принятия решений по проблемам химической и биологической безопасности.

4. Обсуждение: актуальные задачи развития АСИО в области разработки и применения химических веществ и материалов

Чрезвычайное разнообразие и многоаспектность объектов и понятий химии и наук о материалах и ассоциированных областей (в первую очередь, связанных с применением химических веществ и материалов и их воздействием на окружающую среду и здоровье), не позволяло до настоящего времени реализовать интегральный подход к созданию стандартизированной понятийно-терминологической системообразующей базы данных в этой предметной области. Следствием являются трудности как при поиске информации, так и при содержательной переработке и переводе научно-технических документов. Отсутствие системообразующей понятийно-терминологической базы данных является в настоящее время и основным препятствием для автоматизации содержательной переработки электронных научно-технических документов в области химии и наук о материалах при создании/пополнении баз данных в информационных центрах.

Развитие проекта [1, 2] предполагает разработку системы взаимосвязанных понятийно-терминологических справочников (интегрированной базы данных) на основе СУБД, с совершенствованием разработанных и созданием новых методологических и программно-информационных средств для автоматизации содержательной переработки электронных научно-технических документов в области химии и наук о материалах. Система понятийно-терминологических справочников должна представлять систематизированные, формализованные и унифицированные/ стандартизированные данные в области химии и наук о материалах и ассоциированных областей: ядерную базу химических соединений, авто-

матизированные словари терминов, понятий, синонимов, сокращений, дескрипторов (в том числе подструктурных фрагментов и их наименований), указателей, организованные в среде современной СУБД.

Каждый справочник должен содержать информацию по объектам, сгруппированным по определенным признакам. Система справочников должна реализовывать централизованное хранение данных (на служебном сервере хранилища данных и/или локальных АРМ) и унифицированный доступ к данным. Источниками данных должны служить как существующие локальные (для конкретных подразделов/аспектов химии и наук о материалах) тезаурусы и словари ключевых слов в печатном и в электронном виде (файловом), так и вновь формируемые понятийно-терминологические ресурсы с использованием реферативных, полнотекстовых и документально-фактографических баз данных ВИНТИ, РФФИ, SCI, Medline и др., патентных баз данных с акцентом на актуальные направления материальной химии.

Система понятийно-терминологических справочников должна быть интегрирована в пользовательские интерфейсы АРМ экспертов-индексаторов/референтов, АРМ для формирования аналитических отчетов и других информационных продуктов и услуг. В пользовательские интерфейсы АРМ должны быть интегрированы также редакторы графоических объектов (молекулярных структур, технологических схем) и номенклатор химических названий.

Для обеспечения жизненного цикла такой понятийно-терминологической и программно-технологической подсистемы в составе крупного информационного центра для автоматизации содержательной переработки электронных научно-технических документов в области химии и наук о материалах должны быть разработаны специальные методические и программно-технологические средства, реализованные в виде служебных АРМ генерации и ведения системы понятийно-терминологических справочников (базы данных).

В качестве наиболее близкого информационного аналога предлагаемого проекта (по информационным аспектам) является компендиум химической терминологии (<http://www.iupac.org/publications/compendium/index.html>), реализованный в рамках деятельности международной организации IUPAC.

Выводы

1. Описано состояние разработки и практической апробации комплекса программно-технологических и информационных средств для создания адаптивных (развивающихся) систем информационного обеспечения (АСИО) в областях разработки и применения химических веществ и материалов, охарактеризована технология производства /генерации/ различных информационных продуктов и услуг в рамках АСИО.
2. Определена принципиальная задача развития АСИО - создание стандартизированной понятийно-терминологической системообразующей базы данных в рассматриваемой предметной области с учетом ассоциированных аспектов.

Литература

1. *Афонина Н.Н., Бирюков А.Л., Гончаренко Д.И., Зацепин В.М., Иванченко В.А., Мартиросов А.К., Нефедов О.М., Остапчук Р.В.* Развитие банка данных информационных ресурсов ВИНТИ по химии и химической технологии: от структурных данных к материальной химии и технологии. 1. Разработка концепции и исследовательских прототипов регистров химических информационных объектов. Мембраны. 2001, №12, с.38-51.
2. *Афонина Н.Н., Бирюков А.Л., Гончаренко Д.И., Зацепин В.М., Иванченко В.А., Кочанова Н.Н., Мартиросов А.К., Нефедов О.М., Остапчук Р.В., Сафронова О.С., Федоровская М.А.* Развитие банка данных информационных ресурсов ВИНТИ по химии и химической технологии: от структурных данных к материальной химии и технологии. 2. Программно-технологические решения регистров химических соединений и реакций. Мембраны. 2002, №16, с.39-52.
3. *Гончаренко Д.И., Афонина Н.Н., Бирюков А.Л., Зацепин В.М., Иванченко В.А., Круглова Н.А., Мартиросов А.К.* Разработка автоматизированного рабочего места для содержательной переработки научно-технической и патентной информации и формирования массивов реферативно-библиографических данных с включением химико-структурных данных. Материалы 6-й Междунар. конф. «НТИ-2002», 16-18 октября 2002 г. М.: ВИНТИ РАН, 2002, с.116.
4. *Остапчук Р.В., Бирюков А.Л., Гончаренко Д.И., Зацепин В.М., Иванченко В.А., Мартиросов А.К.* Автоматизированная обработка первичных и вторичных источников информации в области химии и химической технологии. Материалы 6-й Междунар. конф. «НТИ-2002», 16-18 октября 2002 г. М.: ВИНТИ РАН, 2002, с.271.
5. *Мартиросов А.К., Бирюков А.Л., Гончаренко Д.И., Зацепин В.М., Иванченко В.А., Остапчук Р.В.* Разработка автоматизированной системы регистрации и классификации химико-структурных данных с использованием формул Маркуша. Материалы 6-й Междунар. конф. «НТИ-2002», 16-18 октября 2002 г. М.: ВИНТИ РАН, 2002, с.241-242.
6. *Сафронова О.С., Бирюков А.Л., Гончаренко Д.И., Зацепин В.М., Иванченко В.А., Мартиросов А.К., Остапчук Р.В.* Разработка системы формализации и регистрации информационных объектов в интегральной многоаспектной базе данных по химии и химической технологии. Материалы 6-й Междунар. конф. «НТИ-2002», 16-18 октября 2002 г. М.: ВИНТИ РАН, 2002, с.303.
7. *Афонина Н.Н., Бирюков А.Л., Гончаренко Д.И., Зацепин В.М., Иванченко В.А., Мартиросов А.К., Нефедов О.М., Остапчук Р.В.* К вопросу о создании национального банка – регистра химических соединений, смесей, химических реакций и процессов. Материалы 5 Междунар. конф. «НТИ-2000», 22-24 ноября 2000 г. М., ВИНТИ РАН. с.51-52.
8. Международная терминология по мембранам и мембранным процессам. Рекомендации IUPAC 1996г. Мембраны. 1999, №2, с.24-37.
9. *Первов А.Г., Хаханов С.А., Дудкин Е.В.* Получение деионизованной воды заданного качества путем комбинации систем обратного осмоса и ионного обмена. Мембраны. 2001, №11, с.3-11.
10. *Кафаров В.В., Мешалкин В.П.* Анализ и синтез химико-технологических систем. –М.: Химия, 1991. - 432с.
11. *Банников А.И., Иванченко В.А., Зацепин В.М.* Проблемы классификации и токсикологической оценки химикатов при реализации Конвенции о запрещении разработки, производства, накопления и применения химического оружия, и о его уничтожении. Сб. «Федеральные и региональные проблемы уничтожения химического оружия» М.: ВИНТИ РАН, 2003, №4, с.16-23.
12. *Зацепин В.М., Иванченко В.А., Левчук М.Н.* Об информационном обеспечении проблем химической и биологической безопасности. Сб. «Федеральные и региональные проблемы уничтожения химического оружия». М.: ВИНТИ РАН. – 2003, №3, 141-144.
13. *Афонина Н.Н., Банников А.И., Бирюков А.Л., Гончаренко Д.И., Зацепин В.М., Иванченко В.А., Левчук М.Н.* Опыт информационного обеспечения проблем химической и биологической безопасности. Материалы 6-й Междунар. конф. «НТИ-2002», 16-18 октября 2002 г. М.: ВИНТИ РАН, 2002, с.45.