

**ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА
ДЛЯ ПОСТРОЕНИЯ БАЗЫ ДАННЫХ
ПО ЛЕКАРСТВЕННЫМ СУБСТАНЦИЯМ,
СНАБЖЕННЫМ ТРАНСПОРТНОЙ СИСТЕМОЙ ДОСТАВКИ**

**О.А. Угольникова, Ю.А. Демич,
В.Ф. Корнюшко, А.В. Лисица**

*ГОУ ВПО «Московская государственная академия тонкой
химической технологии им. М.В. Ломоносова»; ГУ НИИ
биомедицинской химии им. В.Н. Ореховича РАМН, г. Москва*

Рецензент д-р техн. наук, профессор Л.А. Бахвалов

Ключевые слова и фразы: доставка лекарств; информационный менеджмент; проектирование базы данных; семантический анализ; Drug Delivery; PubMed.

Аннотация: Рассмотрена методология информационного менеджмента для построения реляционной базы данных, предназначенной для лекарственных композиций, снабженных транспортными системами доставки и сформированной на основе литературных ссылок из электронных библиотек PubMed и PubChem.

В настоящее время расширение арсенала лекарственных средств для борьбы с основными группами болезней человека остается важнейшей задачей практической медицины и имеет большое социальное значение. При этом проблема разработки систем доставки лекарств является актуальной и быстро развивающейся областью биотехнологии. В основе данного направления лежит комплекс технологических подходов, позволяющих встраивать молекулы лекарственного вещества в состав носителя. При этом изменяется фармакокинетика и фармакодинамика препарата. Такого рода модификации направлены на повышение эффективности действия и снижение побочных эффектов лекарственного вещества как за счет увеличения биодоступности, так и за счет избирательной концентрации вещества в пораженных тканях и органах.

Угольникова О.А. – ассистент кафедры информационных технологий, e-mail: ugolek84@gmail.com; Корнюшко В.Ф. – доктор технических наук, кафедра информационных технологий, Московская государственная академия тонкой химической технологии им. М.В. Ломоносова; Демич Ю.А. – кандидат биологических наук, научный сотрудник лаборатории биоинформационных технологий; Лисица А.В. – доктор биологических наук, заместитель директора по научной работе лаборатории биоинформационных технологий, ГУ НИИ биомедицинской химии им. В.Н. Ореховича РАМН, г. Москва.

Высокий уровень активности научно-исследовательских коллективов, работающих в области создания систем транспорта лекарств, отражается в значительном количестве опубликованных работ. В биомедицинской электронной библиотеке PubMed за последние 20 лет размещено более 500 тыс. публикаций, непосредственно описывающих создание и испытание систем транспорта, и более 1,5 млн публикаций, имеющих косвенное отношение к разработке систем транспорта и посвященных биологическим свойствам лекарственных композиций [1]. Обеспечение химиков-технологов достоверной информацией о свойствах и технологиях получения современных веществ является необходимым условием развития современной промышленности. В настоящее время информационное обеспечение специалистов невозможно без использования специализированных баз данных.

Для проектирования базы данных были применены технологии информационного менеджмента, используемые для проектирования информационных систем, а именно: методика функционального проектирования IDEF0 и методика построения моделей потоков данных DFD.

Вербальная модель предметной области может быть составлена в произвольной форме – в виде неструктурированного текста, набора отдельных высказываний, табличной или графической форме. Это может быть, к примеру, вербальная модель, включающая данные, необходимые для построения базы данных:

- перечень ключевых терминов;
- электронная библиотека PubMed;
- электронная библиотека PubChem;
- целевая выборка;
- фоновая выборка;
- перечень всех специфических терминов;
- контролируемый словарь лекарственных соединений;
- focus_terms.csv;
- names.csv;
- база данных Drug Delivery;
- обучающая выборка,

а также функции:

- составить перечень ключевых слов;
- создать целевую выборку;
- создать фоновую выборку;
- формировать перечень всех специфических терминов;
- выбрать химические соединения;
- подготовить файлы для базы данных;
- загрузить файлы в базу данных;
- обработать базу данных;
- обучить программу при помощи созданной выборки,

и средства:

- эксперт;
- программное обеспечение;
- пользователь.

Следующим этапом является формализация вербальной модели. Функциональная модель для прогнозирования встраивания лекарственных соединений в транспортные системы доставки разработана в соответствии с методологией IDEF0. Эта методология построения функциональных моделей основана на подходе структурного анализа и проектирования систем SADT (Structured Analysis & Design Technique) и графическом способе описания и моделирования систем. В настоящее время это один из наиболее часто используемых в мире стандартов в области CASE-технологий.

В Российской Федерации методология функционального моделирования утверждена Госстандартом России (Р 50.1.028–2001) и имеет статус рекомендаций по стандартизации. Контекстная диаграмма функциональной модели (рис. 1) представляет моделируемую систему в виде «черного ящика».

Декомпозиция контекстной диаграммы позволяет уточнить модель с требуемой степенью детализации, при этом обеспечивается соответствие граничных стрелок на родительской и дочерней диаграммах при помощи ICOM-кодирования стрелок (I-Input, C-Control, O-Output, M-Mechanism). На рис. 2 показана декомпозиция первого уровня блока A-0 контекстной диаграммы.

Следующий этап обеспечивает переход от моделирования функций к моделированию данных. Модель потоков данных, помимо внешних по отношению к моделируемой системе сущностей, включает в себя как наименования функций (процессов), так и абстрактные накопители данных (документы, базы данных и т.д.). Диаграмма потоков данных DFD (Data Flow Diagram), приведенная на рис. 3, построена с использованием нотации Гейна–Сарсона.

На следующем этапе решается задача сбора и первичного анализа литературных данных и выборки литературных источников, содержащих информацию по соединениям, участвующим в липидной системе доставки лекарств. В базе данных массив данных представляется в виде взаимосвязей

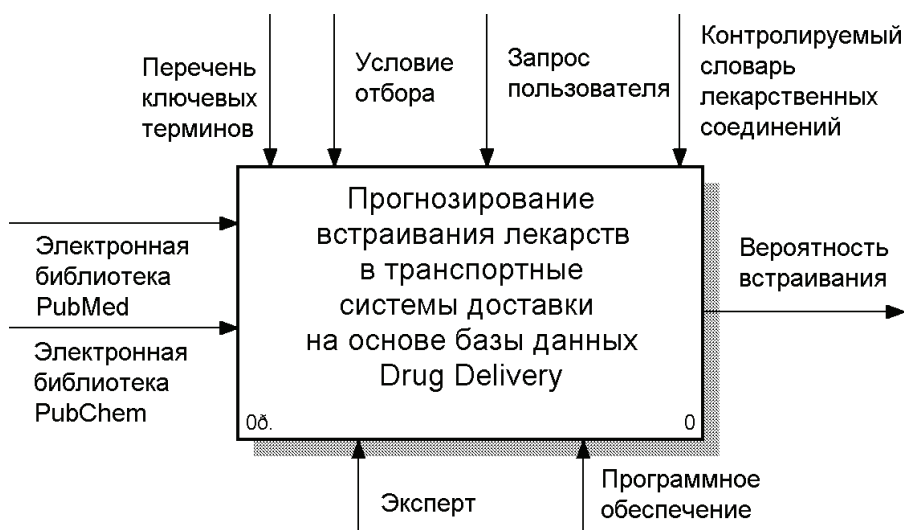


Рис. 1. Контекстная диаграмма A-0 функциональной модели

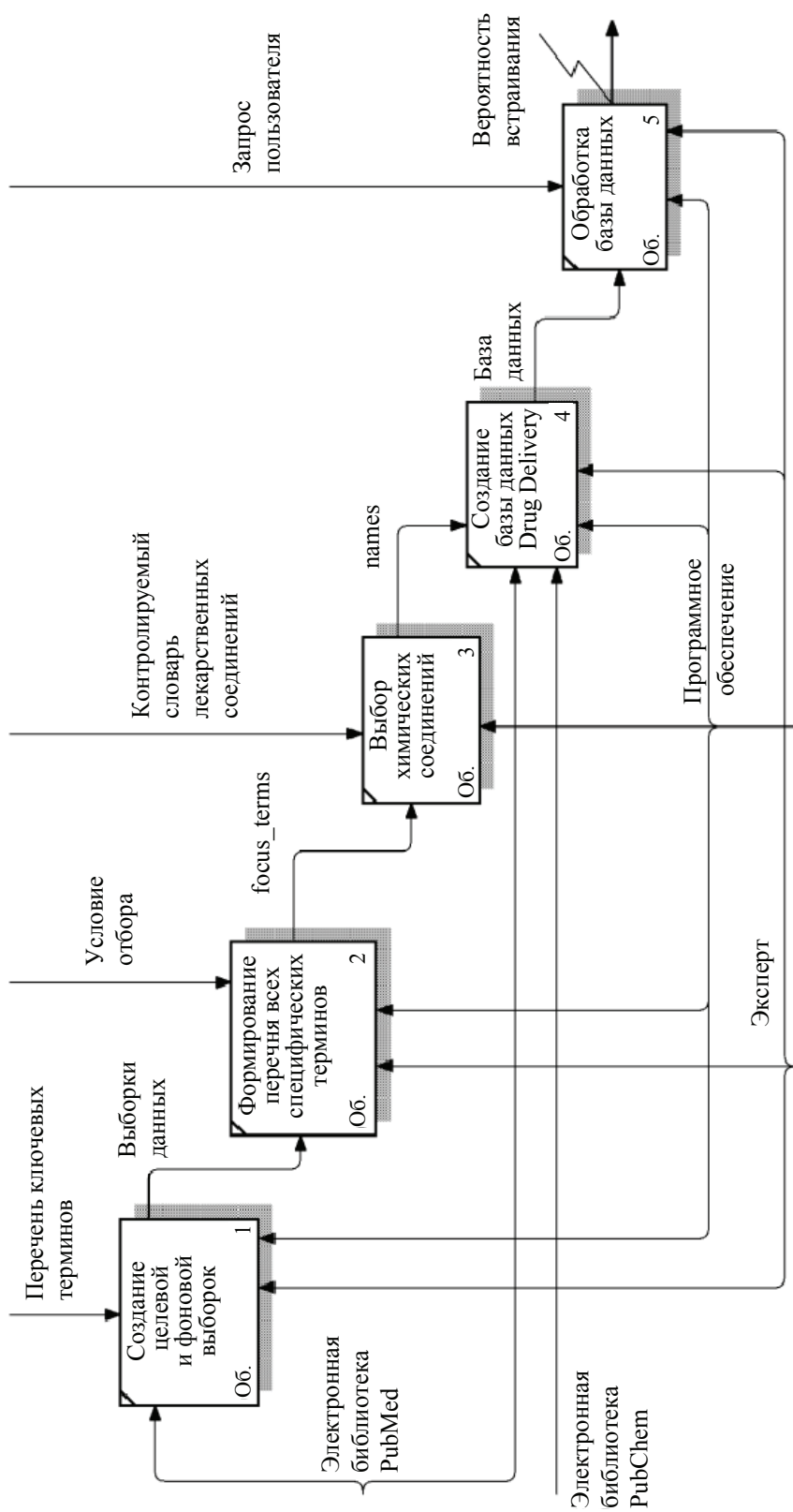


Рис. 2. Декомпозиция первого уровня A0-функциональной модели

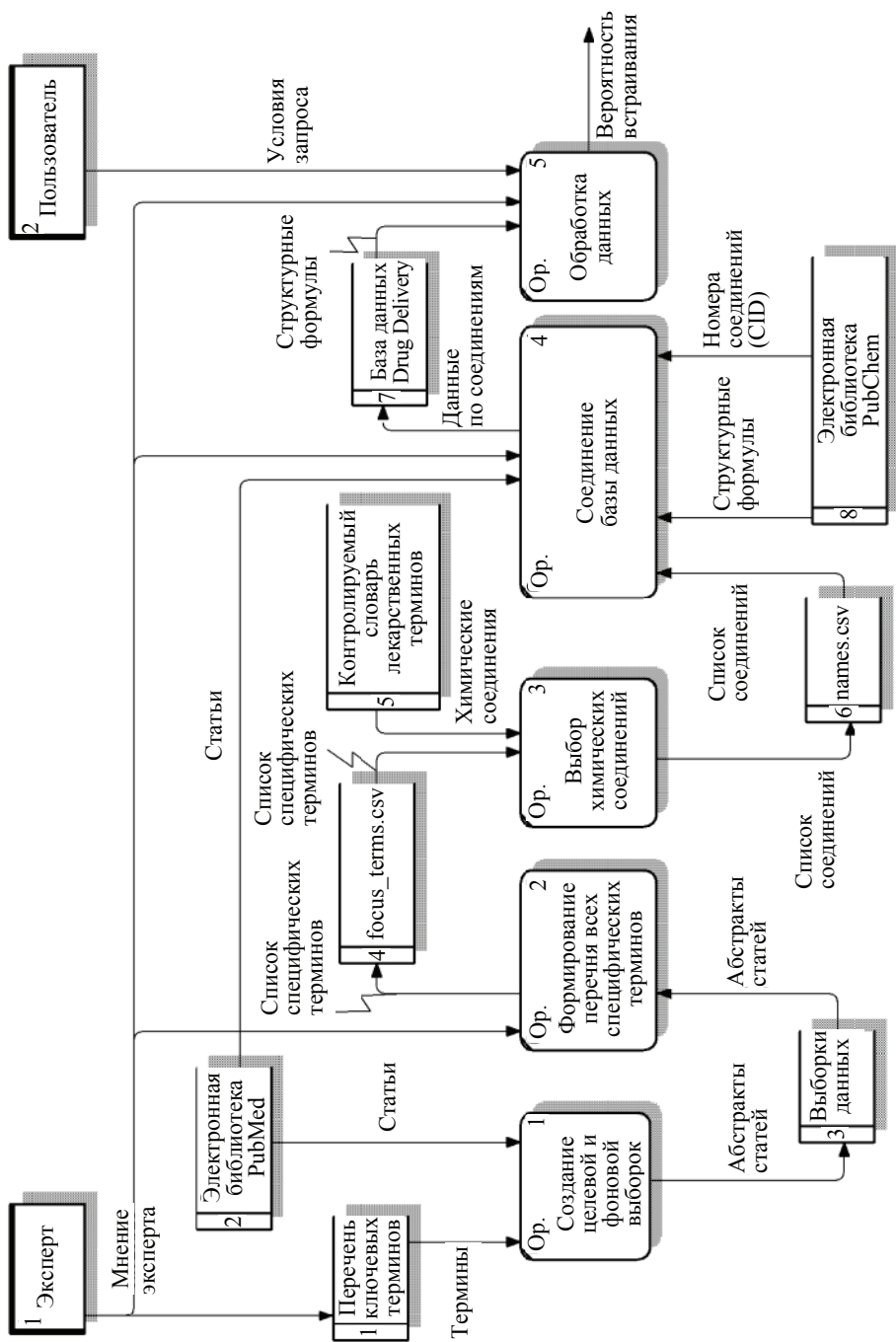


Рис. 3. Диаграмма потоков данных

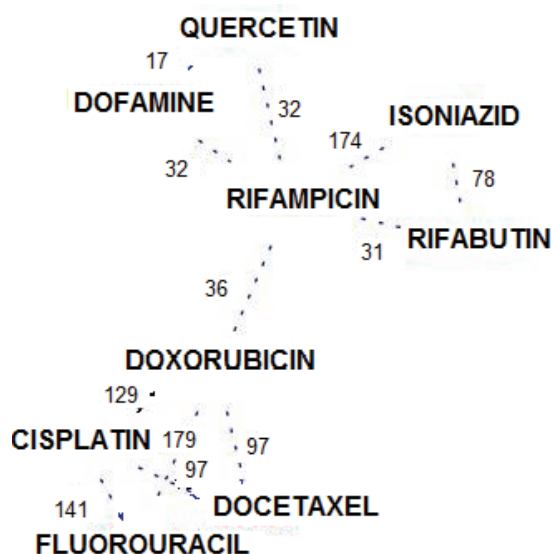


Рис. 4. Фрагмент одного из кластеров

между информационными объектами, которыми являются активные и лекарственные соединения. В основе базы данных лежит представление о том, что объективно существующие взаимосвязи между информационными объектами могут быть выявлены путем статистической обработки достаточного количества субъективных обращений к этим объектам. Для этого были использованы процедуры библиометрического анализа, позволяющие алгоритмическим способом сопоставлять объекты, имеющие индивидуальные обозначения в текстах научных статей [3]. В качестве таких объектов в работе были рассмотрены наименования 2358 химических соединений, используемых в фармакологии. Ассоциативные связи между объектами устанавливали путем сопоставления библиографических списков, полученных для каждого соединения в результате поиска соответствующих наименований в реферативной библиотеке PubMed [1]. Для визуализации результатов обработки использовали программу построения сетевых диаграмм GVedit [4]. Анализ полученных данных позволил найти обособленные кластеры, в которые вошли соединения, характеризующиеся общностью химической структуры или сходными свойствами. На рис. 4 приведен пример сетевого отображения одного из кластеров (цифрами обозначено количество родственных публикаций между действующими веществами липосомальных лекарственных препаратов).

Из рис. 4 видно, что исследуемые соединения сформировали две основные группы: химиотерапевтические антинеопластические агенты (доксорубицин, цисплатин, доцетаксел и фтороурацил) и антибактериальные препараты (в основном ингибиторы нуклеиновых кислот), отдельно выделен кверцетин, обладающий ангиопротекторными и антиоксидантными свойствами, а также дофамин – нейромедиатор, используемый для лечения болезни Паркинсона.

Далее отобранные литературные данные, объекты исследования (2358 соединений), структурные формулы и их идентификаторы (CID) были загружены в базу данных Drug Delivery.

На следующем этапе исследований данные из базы данных Drug Delivery будут использованы для решения задачи прогнозирования встраивания лекарственных средств в транспортные системы доставки. Для выполнения работы необходимо обработать полученные данные для расчета эффективности встраивания лекарственных веществ в системы доставки с использованием компьютерной программы PASS [5].

Список литературы

1. PubMed [Электронный ресурс] : поисковая система, разработ. для работы с базами данных MEDLINE, поддерживаемыми Нац. мед. б-кой США. – Режим доступа : www.pubmed.org. – Загл. с экрана.
2. PubChem [Электронный ресурс] : публич. база данных Нац. мед. б-ки США. – Режим доступа : <http://pubchem.ncbi.nlm.nih.gov/>. – Загл. с экрана.
3. Stapley, B.J. Biobibliometrics information retrieval and visualization from co-occurrence of gene names in Medline abstracts / B.J. Stapley, G. Benoit // Pac symp Biocomput. – 2000. – P. 529–540.
4. Graphviz – Graph Visualization Software [Электронный ресурс]. – Режим доступа : www.graphviz.org. – Загл. с экрана.
5. Лисица, А.В. База знаний по цитохромам P450: разработка и применение : автореф. дис. ... д-ра биол. наук : 03.00.28 / А.В. Лисица. – Москва, 2007. – 44 с.

Information Analysis System for Building a Database on Medicinal Substances Supplied with Transport Systems of Delivery

O.A. Ugolnikova, Yu.A. Demich, V.F. Korniyushko, A.V. Lisitsa

Moscow State Academy of Fine Chemical Technology named after M.V. Lomonosov; Institute of Biomedical Chemistry VN RAMS, Moscow

Key words and phrases: database design; the delivery of drugs; Drug Delivery; information management; PubMed; semantic analysis.

Abstract: The paper considers the methodology for information management to build a relational database designed for pharmaceutical compositions supplied with transport delivery systems formed on the basis of references of the electronic libraries PubMed and PubChem.

© О.А. Угольников, Ю.А. Демич,
В.Ф. Корнюшко, А.В. Лисица, 2010