

УДК: 004.825

## Концептуальное моделирование в исследовании биомедицинских данных

Богатырев М.Ю.\*<sup>1</sup>, Вакурин В.С.\*\*<sup>1</sup>

<sup>1</sup>Тульский государственный университет, Тула, 300600, Россия

**Аннотация.** Статья посвящена разработке эффективного инструмента анализа данных, содержащихся в текстах научных публикаций, основанного на концептуальном моделировании. Рассматривается принцип концептуального моделирования и его основные модели. Приводятся примеры практического применения технологии концептуального моделирования на информационных ресурсах системы PubMed. Данная технология позволяет использовать ресурсы PubMed для создания локальных тематических баз знаний.

**Ключевые слова:** концепт, отношение, понятие, формальный контекст, концептуальный граф, решетка понятий, импликация, PubMed.

### ВВЕДЕНИЕ

Современная биология характеризуется непрерывным ростом числа публикаций. В некоторых направлениях, таких как исследование генома и биоинформатика, этот рост носит экспоненциальный характер [1]. Информационные системы каталогизации публикаций позволяют оперативно получать сведения о них, однако, ввиду огромного числа ссылок, данная информация может оказаться бесполезной. Так запрос к известной системе PubMed [2] в виде фразы «gene expression» приводит к получению 907321-й ссылки, и к моменту выхода данной статьи это число, несомненно, возрастет. Очевидно, что прочитать такое количество публикаций невозможно, и необходима дальнейшая автоматизация обработки полученных ссылок с целью извлечения из соответствующих им текстов полезной информации.

В современном анализе данных извлечение новой нетривиальной информации из текстовых данных называется Text Mining [3]. Характерно, что один из первых результатов применения методов Text Mining к текстам публикаций был получен в медицине: в работе [4] предложен алгоритм установления связей между терминами, встречающимися в разных медицинских публикациях, посвященных мигрени.

В дальнейшем применительно к текстам биомедицинской тематики сформировалось целое направление: Biomedical Text Mining [5, 6] – далее BTM. Состояние исследований в данном направлении отражает обзор [7]. К российским работам в данном направлении следует отнести работу [8], в которой методы Text Mining применены в решении задачи восстановления генных сетей.

Среди задач BTM наиболее значимыми являются задачи *извлечения фактов и построения таксономий*.

---

\*okkambo@mail.ru

\*\*vakourinv1@yandex.ru

Извлечение фактов из текста является общей задачей Text Mining, имеющей разные формулировки и методы решения в зависимости от того, что считается фактом. Методами ВТМ из текстов извлекаются т. н. *именованные сущности* [9], к которым можно отнести названия генов, названия медицинских препаратов, названия заболеваний и т. п. Простейшим фактом считается наличие именованных сущностей в тексте. Более сложный факт – это, например, факт применения определенного препарата при лечении определенного заболевания. Здесь необходимо установить связь – отношение между найденными в тексте именованными сущностями.

Традиционным подходом к решению этой задачи был статистический подход, когда отношения диагностировались по частоте встречаемости слов и словосочетаний. Семантическая составляющая полученных решений в этом случае остается неопределенной. Другие, более современные, методы построения отношений, используют лингвистическую информацию – результаты морфологического и синтаксического анализа текстов. Настоящая работа относится к данному направлению.

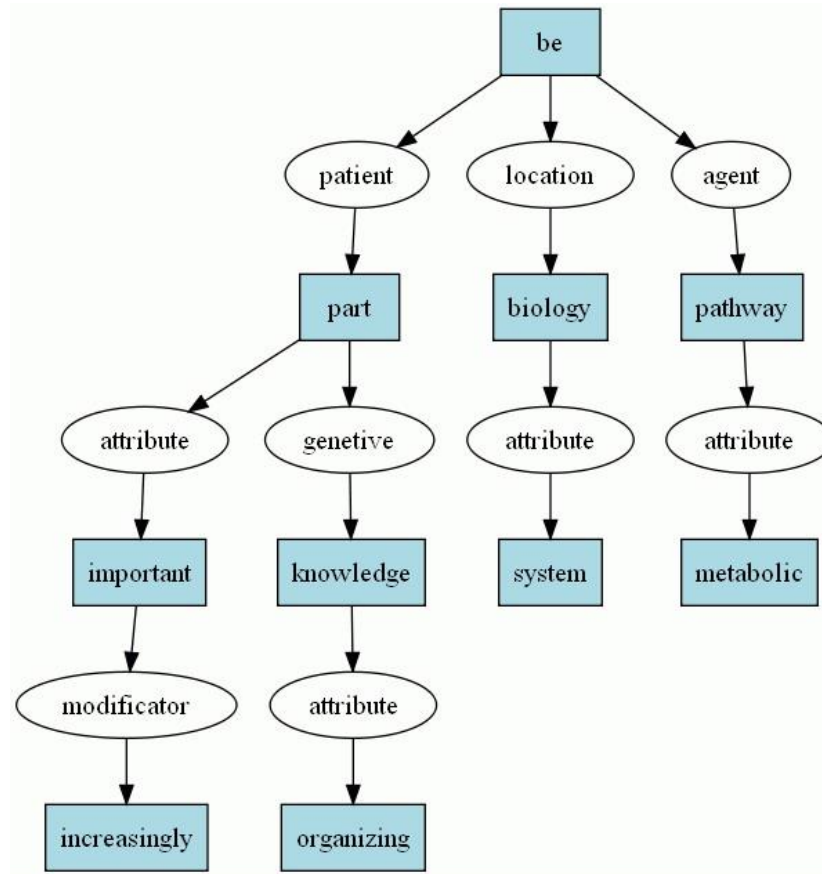
Построение таксономий на множествах текстов – это также известная задача Text Mining [3]. Таксономии являются основой *рубрицирования* текстов – создания структур, отражающих их тематику. На больших ресурсах биомедицинских текстов решение этой задачи принципиально важно: рубрицирование текстов публикаций позволяет определить взаимосвязи между ними и сформировать используемое в текстах множество терминов. Здесь применяются структурные модели, среди которых концептуальные модели играют важную роль. Наиболее масштабной концептуальной моделью, отражающей в целом предметную область множества текстов, является *онтология* [10, 11]. Упомянутая выше система PubMed использует онтологию.

В данной работе рассматривается новая технология ВТМ, основанная на применении концептуального моделирования. В технологии используются две разномасштабных модели: концептуальные графы и решетки понятий. Технология позволяет решать задачи обоих типов: извлечение фактов и построение таксономий.

## ПРИНЦИП КОНЦЕПТУАЛЬНОГО МОДЕЛИРОВАНИЯ

Концептуальное моделирование является достаточно широким направлением в моделировании. Концептуальная модель в общем виде представляет собой множество объектов, связанных отношениями. Эти объекты, называемые «концепт», «концепция», «понятие», могут иметь различную природу. Концептуальная модель инвариантна к природе данных, что позволяет применять концептуальные модели в самых разных областях [12–14]. Вместе с тем, применение концептуального моделирования в конкретных областях требует настройки концептуальных моделей на предметную область.

Элементарной концептуальной моделью является *концептуальный граф* [15]. Концептуальный граф — это двудольный граф, который моделирует семантику отдельного предложения в виде *концептов* и *отношений*. Концептами могут быть отдельные слова предложения текста, отношения строятся путем обработки всего предложения с возможным привлечением внешних ресурсов, например, словарей. Один из способов построения концептуальных графов описан в работе [16]. На рис. 1 показан концептуальный граф предложения «Metabolic pathways are an increasingly important part of organizing knowledge in systems biology», взятого из публикации [17], описывающей одну из технологий ВТМ извлечения фактов.



**Рис. 1.** Концептуальный граф предложения «Metabolic pathways are an increasingly important part of organizing knowledge in systems biology». Концепты изображаются прямоугольниками, отношения – эллипсами.

Число концептуальных графов совпадает с числом предложений текста и может быть весьма велико. При этом, как и некоторые предложения, концептуальные графы могут не содержать полезной информации. Поэтому множество концептуальных графов заменяется агрегированной семантической моделью – *формальным контекстом*. Формальный контекст и далее – решетка понятий являются моделями, используемыми в направлении Data Mining, известном как *формальный анализ понятий* [18].

*Формальный контекст*  $\mathbf{K} = (G, M, I)$  задает связь между множествами объектов  $G$  и принадлежащих им атрибутов  $M$ . Связь определяется отношением  $I \subseteq G \times M$ , которое задается матрицей контекста  $\mathbf{K} = \{k_{i,j}\}$ , ненулевые элементы которой фиксируют факты принадлежности атрибута  $m_j \in M$  объекту  $g_i \in G$ .

Формальный контекст является абстрактной моделью, реализация которой на текстовых данных является сложной нетривиальной задачей. Построение формальных контекстов упрощается на множестве концептуальных графов, поскольку концептуальные графы задают отношения «объект – атрибут» посредством определенных концептуальных отношений – элементов графов. Так, на рис. 1 отношения «attribute» и «genitive» однозначно определяют отношения «объект – атрибут»: например, концепт «metabolic» является атрибутом концепта «pathway» задавая отношение принадлежности. Другие концептуальные отношения также могут задавать отношения «объект – атрибут» – это зависит от смысла конкретных предложений.

## Решетка понятий

Формальный контекст является источником данных для построения содержащихся в нем «понятий». Как и сам контекст, «понятие» является формальным объектом. Тем не менее, существующая на каждом контексте система понятий, образующая т. н. *решетку понятий*, отражает основные взаимосвязи между данными, содержащимися в текстах, на которых построен формальный контекст. При этом при построении контекста и решетки понятий происходит фильтрация несущественных, чисто лексических элементов текста, и полученная в итоге концептуальная структура отражает реальные, возможно, неочевидные, скрытые взаимосвязи между понятиями, выделенными из текста. В этом состоит суть применения методов Text Mining к текстам.

Рассмотрим построение решеток понятий на контекстах. Используемые в контексте множества объектов  $G$  и атрибутов  $M$  должны быть частично упорядочены некоторыми отношениями, которые мы обозначим  $\sqsubseteq$  и  $\in$ , соответственно:  $G = (G, \sqsubseteq)$ ,  $M = (M, \in)$ . В контексте  $\mathbf{K} = (G, M, I)$  связи между объектами и атрибутами определяются следующим образом. Для подмножеств  $A \subseteq G$  и  $B \subseteq M$  объектов и атрибутов задаются отображения (функции)  $A' : A \rightarrow B$  и  $B' : B \rightarrow A$  со следующими свойствами:  $A' := \{m \in M \mid \langle g, m \rangle \in I \text{ для всех } g \in A\}$ ,  $B' := \{g \in G \mid \langle g, m \rangle \in I \text{ для всех } m \in B\}$ . Пара множеств  $(A, B)$ , таких, что,  $A' = B$ ,  $B' = A$  называется *формальным понятием* контекста  $\mathbf{K}$ .

Множества  $A$  и  $B$  замкнуты в силу композиции отображений:  $A'' = A$ ,  $B'' = B$ . Множество  $A$  образует *объем* формального понятия  $(A, B)$ , а множество  $B$  – его *содержание*. Отношения частичного порядка  $\sqsubseteq$ ,  $\in$  на множествах  $G$  и  $M$  индуцируют отношение частичного порядка  $\leq$  на множестве понятий.

Если для понятий  $(A_1, B_1)$  и  $(A_2, B_2)$   $A_1 \subseteq A_2$ , что эквивалентно  $B_2 \subseteq B_1$ , то  $(A_1, B_1) \leq (A_2, B_2)$ . В этом случае логично считать понятие  $(A_1, B_1)$  *менее общим*, чем понятие  $(A_2, B_2)$ . Согласно основной теореме анализа формальных понятий [18], частично упорядоченное по вложению объемов множество формальных понятий контекста  $\mathbf{K}$  образует математический объект – *решетку*, которая называется *решетка понятий*.

Решетка понятий может быть визуализирована в виде графического объекта, отражающего иерархию взаимосвязей понятий, и служащего средством навигации среди объектов – понятий.

Инструментом извлечения знаний на решетках понятий являются методы Data Mining, использующие модели в виде *импликаций*, *функциональных зависимостей* и *ассоциативных правил* [19].

Импликации  $X \rightarrow Y$  на подмножествах признаков  $X, Y \subseteq M$  имеют место, если  $X' \subseteq Y'$ , т. е. каждый объект, обладающий всеми признаками из множества  $X$ , также обладает всеми признаками из множества  $Y$ . Каждая решетка понятий порождает систему импликаций. Система импликаций позволяет исследовать взаимосвязи между извлеченными из текстов понятиями.

К л а с с и ф и к а ц и я            т е к с т о в            в ы п о л н я е т с я  
п у т е м            с о п о с т а в л е н и я            т е к с т а            з а п р о с а  
п о н я т и я м            р е ш е т к и .            Т е к с т            з а п р о с а  
з а м е н я е т с я            м н о ж е с т в о м            п о н я т и й ,  
и з в л е ч е н н ы х            и з            н е г о            п р и            п о м о щ и  
к о н ц е п т у а л ь н ы х            г р а ф о в .            Д л я            к а ж д о г о

понятия ищется «минимальный» узел решетки, соответствующий минимальному уровню ее иерархии – теме понятия. Если такой узел найден, то понятию сопоставляются ресурсы системы технической поддержки. В результате классификации тексту запроса может быть сопоставлено множество узлов решетки понятий и, соответственно, несколько документов, содержащих относящуюся к тематике запроса информацию.

### **РЕАЛИЗАЦИЯ И ПРИМЕНЕНИЕ ТЕХНОЛОГИИ КОНЦЕПТУАЛЬНОГО МОДЕЛИРОВАНИЯ**

Описанная выше технология концептуального моделирования реализована в пилотном проекте информационной системы SemText, предназначенной для извлечения знаний из текстовых данных.

Структурная схема модулей, входящих в систему, показана на рисунке 2. Система имеет собственную базу данных, в которой хранятся информационные ресурсы системы: словари, лексико-грамматические шаблоны, данные, используемые в режиме обучения системы. Такими данными являются результаты запросов к внешним ресурсам, а также промежуточные данные обработки текстов.



Рис. 2. Структурная схема системы SemText.

Одной из задач, решаемых при работе системы, является идентификация понятий, построенных по формальным контекстам. Понятиям, задаваемым формально как подмножества объектов и их атрибутов, в текстах соответствуют словосочетания. В системе существует специальный модуль обработки словосочетаний, позволяющий их идентифицировать на множестве концептуальных графов.

Модуль поиска системы SemText предназначен для взаимодействия с внешней поисковой системой (PubMed) и для построения ответов на запросы пользователей.

Разработанная технология реализуется на платформе Java 2EE и имеет открытую модульную структуру.

### Экспериментальная апробация системы

Технология концептуального моделирования опробована на экспериментах с системой PubMed [2] в режиме on-line доступа к ресурсам системы.

Для решения задач извлечения фактов и построения таксономий на текстовых данных публикаций используются аннотации этих публикаций. С одной стороны, каждая аннотация кратко и в основных чертах отражает содержание публикации, с другой стороны, применение аннотаций вместо полных текстов публикаций обеспечивает приемлемое быстродействие и использование ресурсов при реализации технологии ВТМ. Аннотации публикаций могут быть получены как результат обращения к системе PubMed с текстовым запросом.

Тексты полученных аннотаций публикаций преобразуются в концептуальные графы, по которым строится контекст публикации. По контекстам полученного множества публикаций строится решетка понятий.

Первоначальная настройка на контекст каждой статьи выполняется по прилагаемым к статьям ключевым словам и словосочетаниям. На этапе формирования формального контекста решетки понятий использовался интерактивный режим отбора объектов и их атрибутов с целью исключения заведомо неинформативных пар и сокращения размеров решетки.

Контексты аннотаций задаются множествами словосочетаний разной длины, определяющих тематику статьи, аннотация которой анализируется. Соответственно, на множестве концептуальных графов, построенных на предложениях аннотаций, решается задача поиска словосочетаний методом, описанным в [13]. Данный метод позволяет находить словосочетания в текстах в случаях, когда образующие их слова не являются соседними, что является главным преимуществом метода.

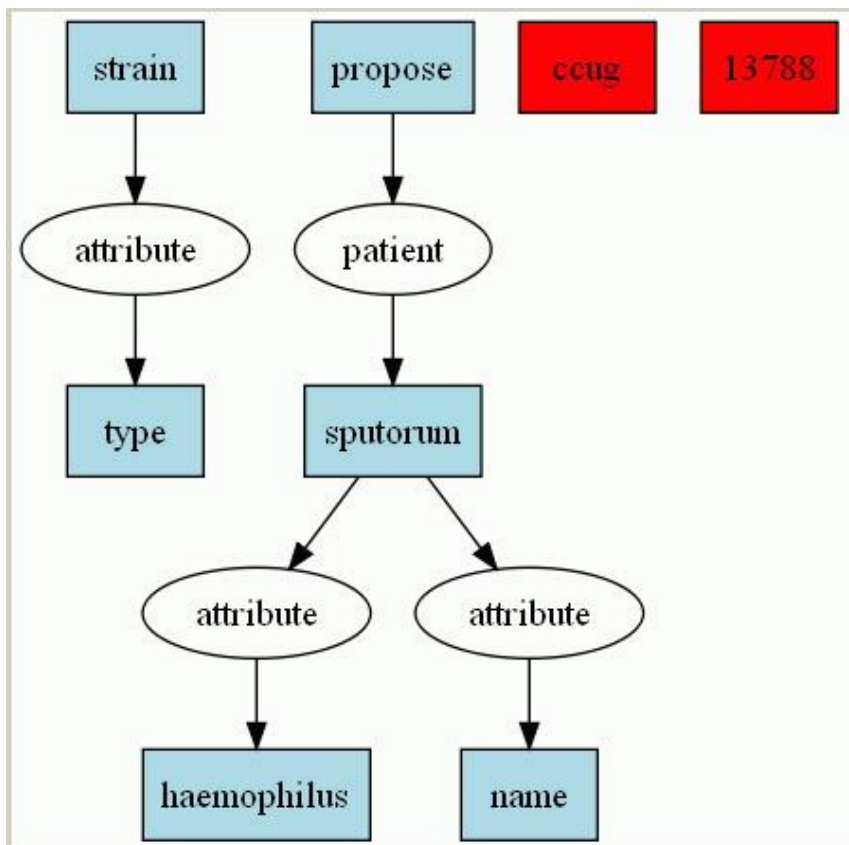
Рассмотрим характерные примеры применения данной технологии.

### **Извлечение фактов. Применение изолированных концептов**

Как отмечено выше, наиболее востребованной задачей ВМТ является нахождение именованных сущностей в тексте. Для решения этой задачи применяются концептуальные графы. При этом используются т. н. *изолированные концепты* – концепты концептуального графа, не имеющие связей в виде отношений с другими концептами. При построении концептуальных графов на текстах естественного языка изолированные концепты возникают тогда, когда алгоритму не хватает данных для построения отношений. Хороший алгоритм не должен создавать изолированные концепты, однако исключить их появление полностью невозможно.

Известной тенденцией является специализация систем обработки текстовых данных по типу и тематике текстов. В этом случае применение концептуальных графов дает хорошую возможность использовать изолированные концепты при обучении системы.

На рис. 3 показан пример изолированного концепта при обработке предложения «The name *Haemophilus sputorum* is proposed, with CCUG 13788(T) (= DSM 24472(T) = NCTC 13537(T)) as the type strain», взятого из аннотации публикации [20]. Текст в скобках считается независимой фразой и обрабатывается отдельно. В данном случае это символно-числовая последовательность, которая, учитывая пробелы, распознается как три изолированных концепта. Концептуальный граф для остальной части предложения изображен на рисунке. Красным цветом показаны изолированные концепты, которые соответствуют аббревиатуре CCUG и номеру 13788, присвоенным бактерии *Haemophilus sputorum*.



**Рис. 3.** Появление изолированных концептов при обработке предложения «The name *Haemophilus sputorum* is proposed, with CCUG 13788(T) (= DSM 24472(T) = NCTC 13537(T)) as the type strain».

В системе SemText изолированные концепты фиксируются и используются для обучения. Обучение приводит к формированию новых типов отношений, которые будут применяться при построении концептуальных графов и исключают появление изолированных концептов.

Так, обнаруженные в тексте изолированные концепты «CCUG» и «13788» соотносятся с классами «аббревиатура» и «число», и далее для этих классов создаются шаблоны, позволяющие в дальнейшем формировать новое отношение «обозначение» и применять существующее отношение «атрибут» при обработке текста. Отношение «обозначение» свяжет концепты «propose» и «CCUG», а отношение «атрибут» – концепты «CCUG» и «13788».

Очевидно, что тексты биомедицинской тематики содержат обозначения, поэтому сформированное новое отношение будет применяться при разборе текстов. При этом нет необходимости настраивать систему на каждое новое обозначение. Класс концептов «аббревиатура» формируется так, что, например, обозначения белковых структур также будут отнесены к аббревиатурам и классифицированы как обозначения.

Анализируя далее граф на рис. 3, заметим, что хотя бактерия *Haemophilus sputorum* – это одно существо, его название раскрыто в виде двух концептов, связанных отношением «атрибут». В результате применения метода извлечения словосочетаний из текстов бактерия *Haemophilus sputorum* идентифицируется как объект *sputorum* с атрибутом *Haemophilus* и может войти в понятие, объем которого определяется всеми *sputorum* – бактериями, а содержание – всеми бактериями, относящимися к классу *sputorum*.



## Построение таксономий

Построение таксономий является одной из самых сложных задач Text Mining, решаемой, как правило, интерактивно, с участием специалиста. Таксономии моделируются при помощи онтологий, для построения которых непосредственно по текстовым данным применяются решетки понятий [21].

В системе SemText при построении таксономий применяется пользовательский интерфейс, позволяющий редактировать данные, необходимые для построения контекстов и решеток понятий на множествах концептуальных графов, поскольку эти множества, как правило, содержат избыточные данные из-за особенностей изложения материала публикаций в аннотациях. Визуализация решеток понятий не используется как элемент интерфейса, но может быть выполнена с помощью внешнего программного обеспечения [22].

Вернемся к примеру на рис. 3 и рассмотрим его окружение в виде публикаций, которые вместе с публикацией [20] получены из системы PubMed по запросу «gene expression». Публикация [20] входит в множество, сформированное PubMed, поскольку словосочетание «gene expression» встречается в тексте публикации. Однако в аннотации данной публикации словосочетания «gene expression» нет, а в целом публикация посвящена исследованию генома бактерии *Haemophilus sputorum*. При этом система PubMed возвращает лишь 3 публикации на запрос «*Haemophilus sputorum*».

Выясним, каким образом данная редкая публикация относится к теме «экспрессия генов» – «gene expression». В тексте аннотации публикации есть словосочетания «gene sequencing» и «expression of IgA1 protease». Эти словосочетания попадают в понятие «gene», объединяющее словосочетания со словом «gene» из других аннотаций. На этих словосочетаниях строится подрешетка понятий, в которой существует импликация «expression» → «sequence», по которой мы и устанавливаем принадлежность публикации теме «экспрессия генов».

Для построения решетки понятий не применялось все множество из 907321-й аннотации. В эксперименте использовались 5 аннотаций, расположенных до данной, и 8 аннотаций после нее. Данный выбор был обусловлен предположением о том, что в системе PubMed имеет место ранжирование выходного списка публикаций, и в списке рядом с исследуемой публикацией находятся «близкие» к ней по содержанию публикации. Это предположение не оправдалось, как и следовало ожидать, в части «близости» публикаций, поскольку количественное совпадение слов в текстах не может служить мерой их семантической близости. Семантическую близость понятий, сформированных на тексте, определяет решетка понятий через систему импликаций, допускаемых решеткой. При обработке концептуальных графов применялась упомянутая выше фильтрация несущественных элементов концептуальных графов, что позволило существенно сократить число обрабатываемых графов.

## ЗАКЛЮЧЕНИЕ. ПЕРСПЕКТИВЫ РАЗВИТИЯ ТЕХНОЛОГИИ КОНЦЕПТУАЛЬНОГО МОДЕЛИРОВАНИЯ БИМЕДИЦИНСКИХ ДАННЫХ

Разработанная технология концептуального моделирования биомедицинских данных представляет собой эффективный инструмент анализа данных, содержащихся в текстах научных публикаций. Практическое значение данной технологии состоит в том, что она позволяет использовать существующие информационные ресурсы для создания локальных тематических баз знаний.

Применение технологии в конкретной предметной области требует ее настройки путем включения в применяемые в технологии словари терминов и понятий из данной области знания. Такое включение выполняется средствами самой технологии путем идентификации характерных словосочетаний, содержащихся в тексте, и образующих понятия предметной области.

Текущая версия SemText обеспечивает сложность вычислений порядка  $O(n^3)$ , где  $n$  – число обрабатываемых предложений. Это связано с тем, что применяемые здесь алгоритмы построения концептуальных графов используют объекты, отражающие грамматические правила языка – грамматические шаблоны. Типичный алгоритм, использующий шаблоны, осуществляет поиск шаблона, совпадающего с обрабатываемой частью предложения – одним или несколькими словами, – и далее выполняет действия, предписываемые шаблоном. Такой метод использования шаблонов требует полного перебора множества шаблонов, причем в многопроходных алгоритмах перебор может быть неоднократным. Поэтому уменьшение порядка сложности вычислений связано с оптимизацией типового алгоритма и является актуальной задачей.

Развитие данной технологии в предметной области биомедицины планируется в следующих направлениях.

1. Разработка пользовательских интерфейсов, ориентированных на более узкие предметные области. Здесь развитие технологии должно выполняться в сотрудничестве со специалистами в конкретной предметной области биомедицины.

2. Углубление интеграции с внешними системами, в частности, с системой PubMed, путем разработки программных интерфейсов поддержки иерархии уровней поиска. В этом случае появляется возможность обрабатывать не только текст данной публикации, но также тексты публикаций, на которые в данной публикации есть ссылки.

3. Оптимизация алгоритмов построения концептуальных моделей методами комбинаторной оптимизации, что позволит обрабатывать большие биомедицинские информационные ресурсы.

Работа частично поддержана РФФИ, грант 11-07-97542.

## СПИСОК ЛИТЕРАТУРЫ

1. Hunter L, Cohen KB. Biomedical language processing: what's beyond PubMed? *Mol. Cell*. 2006. V. 21. P. 589–94.
2. U.S. National Library of Medicine. URL: <http://www.ncbi.nlm.nih.gov/pubmed>. (дата обращения: 15.03.2013).
3. Feldman R., Sanger J. *The Text Mining Handbook*. New York: Cambridge University Press, 2006
4. Swanson D.R., Smalheiser N.R. An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial Intelligence*. 1997. V. 91. P. 183–203.
5. *Text Mining for Biology and Biomedicine*. Eds. Ananiadou S., McNaught J. Artech House Books, 2006.
6. Shatkay H, Craven M. *Biomedical Text Mining*. Cambridge, Massachusetts: MIT Press, 2007.
7. Zweigenbaum P., Demner-Fushman D., Yu H., Cohen K.B. Frontiers of biomedical text mining: current progress. *Briefings in Bioinformatics*. 2007. V. 8. № 5. P. 358–375.
8. Kolchanov N.A., Podkolodnyy N.L., Ivanisenko V.A., Ananko E.A., Demenkov P.S., Yarkova E.E. Technology For Gene Network Reconstruction and Analysis. In: *Proc. of the 2nd Int. Conf. on Math. Biology and Bioinformatics (Pushchino, Russia)*. Moscow, 2008. P. 86–87.
9. Fukuda K., Tsunoda T., Tamura A., Takagi T. Toward information extraction: identifying protein names from biological papers. In: *Pac Symp Biocomput*. 1998. P. 707–718.

10. *Medical informatics: Advances in knowledge management and data mining in biomedicine*. Eds. Chen H., Fuller S., Hersh W.R., Friedman C. Springer-Verlag, 2005. P. 211–236.
11. *National Center for Biomedical Ontology*. URL: <http://www.bioontology.org/> (дата обращения: 15.03.2013).
12. Bogatyrev M.Y., Tuhtin V.V. Conceptual Modelling In Biomedical Knowledge Discovering. In: *Proc. of the 3<sup>rd</sup> Int. Conf. on Mathematical Biology and Bioinformatics (Pushchino, Russia)*. Moscow, 2010. P. 214–215.
13. Bogatyrev M., Kolossoff A. Using Conceptual Graphs for Text Mining in Technical Support Services. In: *Pattern Recognition and Machine Intelligence*. 2011. V. 6744. P. 466–471. (Lecture Notes in Computer Science).
14. Bogatyrev M., Nuriyahmetov V. Application of Conceptual Structures in Requirements Modeling. In: *Proc. of the International Workshop on Concept Discovery in Unstructured Data (CDUD 2011) at the Thirteenth International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing - RSFDGrC 2011*. Moscow, Russia, 2011. P. 11–19.
15. Sowa J.F. *Conceptual Structures: Information Processing in Mind and Machine*. London, UK: Addison-Wesley, 1984.
16. Bogatyrev M.Y., Mitrofanova O.A., Tuhtin V.V. Building Conceptual Graphs for Articles Abstracts in Digital Libraries. In: *Proceedings of the Conceptual Structures Tool Interoperability Workshop (CS-TIW 2009) at 17<sup>th</sup> International Conference on Conceptual Structures (ICCS'09) (Moscow, Russia, July)*. 2009. P. 50–57.
17. Kano Y., Dobson P., Nakanishi M., Tsujii J., Ananiadou S.. PathText: a text mining integrator for biological pathway visualizations. *Bioinformatics*. 2010. V. 26. № 19. P. 2486–2487.
18. *Formal Concept Analysis: Foundations and Applications*. Eds. Bernhard G., Stumme G., Wille R. №. 3626. Springer-Verlag, 2005. (Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence)
19. Han J., Kamber M., Pei J. *Data Mining: Concepts and Techniques*. Third Edition. Morgan Kaufman, 2012.
20. Nørskov-Lauritsen N., Bruun B., Andersen C., Kilian M. Identification of haemolytic *Haemophilus* species isolated from human clinical specimens and description of *Haemophilus sputorum* sp. nov. *Int J Med Microbiol*. 2012. V. 302. № 2. P.78–83.
21. Bendaoud R., Napoli A., Toussaint Y. Formal Concept Analysis: A unified framework for building and refining ontologies. In: *Proc. 16th International Conference on Knowledge Engineering and Knowledge Management – EKAW*. 2008. P. 156–171.
22. Евтушенко С.А. Система анализа данных "CONCEPT EXPLORER". В: Труды 7-ой национальной конференции по искусственному интеллекту КИИ-2000. М.:Физмалит, 2000. С. 127–134

Материал поступил в редакцию 01.04.2013, опубликован 30.06.2013.