

АНАСТАСИЯ ВИКТОРОВНА МОЧАЛОВА
соискатель кафедры прикладной математики и кибернетики
математического факультета, Петрозаводский государственный университет (Петрозаводск, Российская Федерация)
stark345@gmail.com

СОЗДАНИЕ И ПОПОЛНЕНИЕ ТЕРМИНОЛОГИЧЕСКИХ СИСТЕМ С ПОМОЩЬЮ СЕМАНТИЧЕСКОГО АНАЛИЗАТОРА*

Автоматическое создание и пополнение терминологических систем на сегодняшний день является актуальной задачей. Это связано с тем, что подобные системы применяются для решения широкого спектра задач, связанных с анализом текстовой информации. Часто решение таких задач требует постоянного пополнения терминологических систем (терминосистем) актуальной информацией. В работе предлагается способ автоматического пополнения терминосистем с помощью алгоритма наложения семантических цепочек, специальным образом получаемых из набора семантических зависимостей. Эти семантические зависимости строит семантический анализатор по каждому анализируемому предложению входного русскоязычного текста. Определяются условия формирования семантических цепочек, а также предлагается алгоритм их наложения. Преимущество такого способа пополнения терминосистемы состоит в том, что в качестве анализируемых текстов могут использоваться любые русскоязычные тексты, содержащие проверенную, непротиворечивую информацию, а не только энциклопедические и толковые словари, или готовые онтологии, что позволяет вносить в систему информацию из новейших текстовых источников, например из научных статей, книг, докладов, новостных и аналитических обзоров.

Ключевые слова: терминосистемы, пополнение терминосистем, семантические отношения, семантический анализатор

ВВЕДЕНИЕ

В настоящее время уже существует достаточно большое количество различных терминологических систем. Такие системы могут использоваться как самостоятельно (например, в системах автоматических языковых переводов, в диалоговых, вопросно-ответных, поисковых системах и др.), так и в качестве составляющей более высокоуровневых хранилищ информации, например онтологий.

Одна из важнейших проблем, возникающих при использовании терминосистем, заключается в том, что данные, содержащиеся в них, должны соответствовать действительности: окружающий мир постоянно изменяется, что влечет за собой интенсивный рост объема информации. Для того чтобы уже созданные терминологические системы были актуальными, необходимо постоянно пополнять их новыми реальными данными.

На сегодняшний день имеются попытки автоматизировать процесс создания и пополнения терминосистем. Один из самых распространенных способов пополнения терминосистем базируется на использовании словарных определений. Также существует онтологический подход к построению терминосистем.

Однако большинство терминологических систем создается вручную (примерами подобных систем могут служить системы, особенности построения которых описаны в работах [6], [7]), что зачастую вызывает непреодолимые трудно-

сти в силу крайне большого объема обрабатываемых информационных материалов и малого количества специалистов, способных выполнить подобную работу, или же дороговизны оплаты их труда. Поэтому задача автоматизации пополнения терминосистем в настоящее время является актуальной.

СПОСОБЫ АВТОМАТИЧЕСКОГО ПОПОЛНЕНИЯ ТЕРМИНОСИСТЕМ

Один из самых распространенных способов автоматического пополнения терминологических систем базируется на анализе словарных определений. При таком подходе в терминосистему добавляются сущности и тип их связи, найденные с помощью шаблонного поиска в словарном определении. Например, если требуется собрать все возможные виды строительных инструментов, то поиск в словаре производится по шаблону вида «X является разновидностью строительных инструментов, которые...», «X – это строительный инструмент, используемый для...», «X – вид строительного инструмента, применяемый с целью...» и т. д. Подобные шаблоны поиска могут формироваться как вручную, так и автоматически с помощью самообучающихся программ [9]. К примеру, идея автоматического пополнения терминологической системы, являющейся основной составляющей онтологии, подробно описывается в работах [15], [17], [18]. Автор работы [15], оценивая методы автоматического

пополнения терминологических систем в рамках онтологий, утверждает, что изучение результатов, полученных на двух основных направлениях – *Ontology learning from texts* и *Ontology learning from machine-readable dictionaries*, а также собственный опыт привели его к выводу, что на данном этапе наиболее перспективна с точки зрения получения практических результатов технология, основанная на анализе и формализации определений, содержащихся в традиционных энциклопедических и толковых словарях. Автор работы [15] указывает на наличие отечественных работ, ориентированных на пополнение терминологических систем, находящихся в составе онтологий, на основе русскоязычных текстов. Примерами таких работ могут служить работы наших соотечественников А. Е. Ермакова [4], И. А. Минакова [8], В. И. Пекара [14], А. Нечипоренко и А. Русина [13].

Другой способ автоматизации пополнения терминосистем базируется на использовании онтологий. Например, автор работы [1] описывает онтологический подход к построению терминосистем. В работе [3] на примере терминологии логистики описывается способ создания терминосистем с помощью онтологий.

АВТОМАТИЧЕСКОЕ СОЗДАНИЕ И ПОПОЛНЕНИЕ ТЕРМИНОСИСТЕМЫ

Терминология

Терминосистема – упорядоченное множество терминов с зафиксированными отношениями между ними, отражающими отношения между называемыми этими терминами понятиями [2]. Терминосистему можно представить в виде ориентированного графа, узлами которого являются термины, а направленные ребра соответствуют семантическим зависимостям, связывающим эти термины.

Семантическая зависимость – некая универсальная связь, усматриваемая носителем языка в тексте. Эта связь бинарна, то есть она идет от одного семантического узла к другому узлу [16]. Будем говорить, что два различных слова α и β из одного предложения связывает семантическая зависимость с именем R (обозначим $R(\alpha, \beta)$), если между α и β существует некая универсальная бинарная связь.

Семантическая сеть – информационная модель предметной области, имеющая вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги (ребра) задают отношения между ними. Объектами могут быть понятия, события, свойства, процессы [20].

Семантическим анализатором назовем программу, автоматически создающую базу данных, хранящую анализируемый текст в виде семантической сети.

Факт – достоверное знание, истинность которого доказана [5]. Предполагается, что на вход семантическому анализатору подаются тексты, истинность содержания которых не подлежит сомнению, поэтому в качестве фактов предлагается использовать определенные наборы семантических отношений между словами, словосочетаниями и другими синтаксическими единицами этих текстов на русском языке. О том, как именно формируются эти наборы, будет изложено далее. Подобное определение факта предложено в работе [12], где за факт принята семантическая зависимость, связывающая два термина.

Связные семантические отношения

Рассмотрим все возможные варианты соотношений (с точки зрения точного совпадения и несовпадения) между аргументами двух семантических зависимостей $R_1(\alpha_1, \beta_1)$ и $R_2(\alpha_2, \beta_2)$ (см. табл. 1).

В табл. 1 показано, что случай, когда совпадают аргументы более чем одной из четырех рассматриваемых пар, невозможен. Это связано с тем, что семантическая зависимость не может связывать два одинаковых аргумента, так же как не может быть двух зависимостей, связывающих одни и те же аргументы, расположенные в инверсном порядке, или двух различных семантических зависимостей, связывающих одни и те же аргументы. Другими словами, не существует $R(\alpha, \alpha)$, так же как не может существовать $R_1(\alpha_1, \beta_1)$, если есть $R_2(\beta_1, \alpha_1)$ и не может быть таких различных R_1 и R_2 , для которых будет верно как $R_1(\alpha_1, \beta_1)$, так и $R_2(\alpha_1, \beta_1)$ (случай когда $R_1 = R_2$, $\alpha_1 = \alpha_2$ и $\beta_1 = \beta_2$, очевидно, тоже рассматривать бессмысленно). Таким образом, имеет смысл рассматривать лишь те варианты соотношений между аргументами семантических зависимостей $R_1(\alpha_1, \beta_1)$ и $R_2(\alpha_2, \beta_2)$, которые в табл. 1 отмечены как возможные случаи.

Пару семантических отношений $R_1(\alpha_1, \beta_1)$ и $R_2(\alpha_2, \beta_2)$ назовем связной, если выполнено одно из четырех следующих условий:

$$(\alpha_1 \neq \alpha_2) \& (\beta_1 \neq \beta_2) \& (\alpha_1 \neq \beta_2) \& (\beta_1 = \lambda_2) \quad (1)$$

$$(\alpha_1 \neq \alpha_2) \& (\beta_1 \neq \beta_2) \& (\alpha_1 = \beta_2) \& (\beta_1 \neq \lambda_2) \quad (2)$$

$$(\alpha_1 \neq \alpha_2) \& (\beta_1 = \beta_2) \& (\alpha_1 \neq \beta_2) \& (\beta_1 \neq \lambda_2) \quad (3)$$

$$(\alpha_1 = \alpha_2) \& (\beta_1 \neq \beta_2) \& (\alpha_1 \neq \beta_2) \& (\beta_1 \neq \lambda_2) \quad (4)$$

Пример 1: Рассмотрим анализируемое предложение: «Скачет, мчится по пыльной дороге уставший всадник». Для этого предложения будут составлены семантические связи:

- *Основа (всадник, мчаться)* (R1)

- *Основа (всадник, скакать)* (R2)

- *Характеристика действия (скакать, быстро)* (R3)

- *Местоположение (скакать, по дороге)* (R4)

- *Местоположение (мчаться, по дороге)* (R5)

- *Признак (дорога, пыльный)* (R6)

Таблица 1

Соотношение между аргументами семантических зависимостей $R_1(\alpha_1, \beta_1)$ и $R_2(\alpha_2, \beta_2)$

№	Соотн. между α_1 и α_2	Соотн. между β_1 и β_2	Соотн. между α_1 и β_2	Соотн. между β_1 и α_2	Выводы
1	$\alpha_1 \neq \alpha_2$	$\beta_1 \neq \beta_2$	$\alpha_1 \neq \beta_2$	$\beta_1 \neq \alpha_2$	Недостаточно информации
2	$\alpha_1 \neq \alpha_2$	$\beta_1 \neq \beta_2$	$\alpha_1 \neq \beta_2$	$\beta_1 = \alpha_2$	Возможный случай № 1
3	$\alpha_1 \neq \alpha_2$	$\beta_1 \neq \beta_2$	$\alpha_1 = \beta_2$	$\beta_1 \neq \alpha_2$	Возможный случай № 2
4	$\alpha_1 \neq \alpha_2$	$\beta_1 \neq \beta_2$	$\alpha_1 = \beta_2$	$\beta_1 = \alpha_2$	Противоречие
5	$\alpha_1 \neq \alpha_2$	$\beta_1 = \beta_2$	$\alpha_1 \neq \beta_2$	$\beta_1 \neq \alpha_2$	Возможный случай № 3
6	$\alpha_1 \neq \alpha_2$	$\beta_1 = \beta_2$	$\alpha_1 \neq \beta_2$	$\beta_1 = \alpha_2$	Противоречие
7	$\alpha_1 \neq \alpha_2$	$\beta_1 = \beta_2$	$\alpha_1 = \beta_2$	$\beta_1 \neq \alpha_2$	Противоречие
8	$\alpha_1 \neq \alpha_2$	$\beta_1 = \beta_2$	$\alpha_1 = \beta_2$	$\beta_1 = \alpha_2$	Противоречие
9	$\alpha_1 = \alpha_2$	$\beta_1 \neq \beta_2$	$\alpha_1 \neq \beta_2$	$\beta_1 \neq \alpha_2$	Возможный случай № 4
10	$\alpha_1 = \alpha_2$	$\beta_1 \neq \beta_2$	$\alpha_1 \neq \beta_2$	$\beta_1 = \alpha_2$	Противоречие
11	$\alpha_1 = \alpha_2$	$\beta_1 \neq \beta_2$	$\alpha_1 = \beta_2$	$\beta_1 \neq \alpha_2$	Противоречие
12	$\alpha_1 = \alpha_2$	$\beta_1 \neq \beta_2$	$\alpha_1 = \beta_2$	$\beta_1 = \alpha_2$	Противоречие
13	$\alpha_1 = \alpha_2$	$\beta_1 = \beta_2$	$\alpha_1 \neq \beta_2$	$\beta_1 \neq \alpha_2$	Противоречие
14	$\alpha_1 = \alpha_2$	$\beta_1 = \beta_2$	$\alpha_1 \neq \beta_2$	$\beta_1 = \alpha_2$	Противоречие
15	$\alpha_1 = \alpha_2$	$\beta_1 = \beta_2$	$\alpha_1 = \beta_2$	$\beta_1 \neq \alpha_2$	Противоречие
16	$\alpha_1 = \alpha_2$	$\beta_1 = \beta_2$	$\alpha_1 = \beta_2$	$\beta_1 = \alpha_2$	Противоречие

• **Признак (всадник, уставший)** (R7)

В табл. 2 приведены пары семантических зависимостей, сформированные семантическим анализатором по предложению из примера 1. Каждый столбец таблицы содержит пары семантических отношений, для которых выполняются условия (1)–(4) соответственно.

Таблица 2

Пары связанных семантических отношений, классифицированных в соответствии с условиями (1)–(4)

Пары сем. отнош., для которых верно условие (1)	Пары сем. отнош., для которых верно условие (2)	Пары сем. отнош., для которых верно условие (3)	Пары сем. отнош., для которых верно условие (4)
(R1)–(R5) (R4)–(R6)	(R3)–(R2)	(R5)–(R4)	(R1)–(R7) (R1)–(R2)

Семантическая цепочка

Семантической цепочкой назовем такой набор различных семантических связей $R_{\alpha\beta} = \{R_1(\alpha_1, \beta_1), R_2(\alpha_2, \beta_2), \dots, R_n(\alpha_n, \beta_n)\}$, в котором все семантические связи можно упорядочить таким образом, что любое семантическое отношение, не стоящее на первом месте, будет связано хотя бы с одним семантическим отношением, стоящим до него (не обязательно с соседним). При этом α_i , являющееся подлежащим в анализируемом предложении, по которому строится семантическая цепочка, назовем головным словом данной цепочки.

В работе [19] авторы определяют понятие *лексической цепочки* как последовательности слов текста, в которой каждое следующее слово связа-

но некоторым отношением с предшествующими словами цепочки. Такие цепочки могут выходить за рамки одного предложения и проходить через целый текст.

Опираясь на это определение, семантические цепочки можно рассматривать как частный случай представления лексических цепочек, в которых отношения между узлами являются семантическими зависимостями, а сама семантическая цепочка рассматривается в рамках одного предложения. Однако узлы семантической цепочки имеют более высокий уровень в иерархии языковых единиц, так как являются не просто словами, как в случае с лексической цепочкой, а неделимыми смысловыми единицами, которые могут быть представлены группой слов.

Пример 2: Рассмотрим предложение, приведенное в Примере 1. Для этого предложения семантическая цепочка будет представлена следующим множеством семантических связей $R_{\alpha\beta}$:

$$R_{\alpha\beta} = \{$$

Основа (всадник, мчаться)
Основа (всадник, скакать)
Характеристика_действия (скакать, быстро)
Местоположение (скакать, по дороге)
Местоположение (мчаться, по дороге)
Признак (дорога, пыльный)
Признак (всадник, уставший)

$$\},$$

так как все отношения из этого множества можно переставить таким образом, что любое семантическое отношение, не стоящее на первом месте, будет связано хотя бы с одним семантическим отношением, стоящим до него. На рис. 1 нагляд-

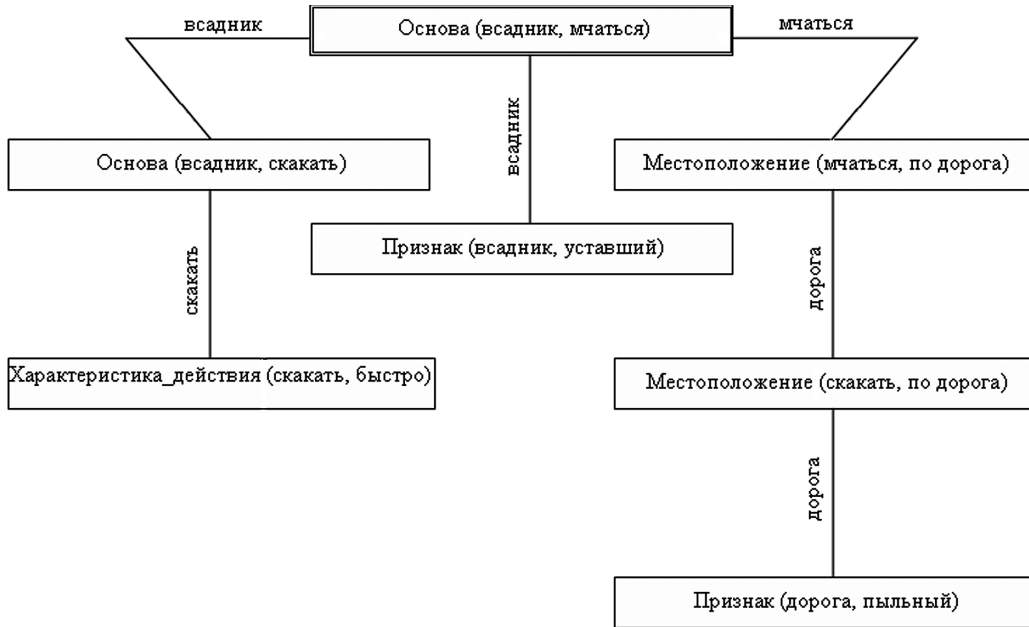


Рис. 1. Пары связанных семантических зависимостей

но представлены пары связанных семантических отношений, переставленные вышеописанным способом.

Очевидно, что все семантические зависимости из правильно построенного предложения будут принадлежать одной семантической цепочке.

Семантической цепочке соответствует ориентированный граф, в котором узлы – это слова, словосочетания или части предложения, являющиеся аргументами семантических зависимостей; вершины этого графа соединяют ребра, направленные от первого аргумента семантической связи ко второму. Ребра графа имеют названия,

однoименные с семантическими связями, связывающими пару вершин данного графа.

На рис. 2 представлен семантический граф, сформированный по вышеописанным правилам и отражающий семантическую цепочку, сформированную по предложению из Примера 1 (головное слово обведено двойной рамкой).

Алгоритм наложения семантических цепочек

Результатом наложения двух семантических цепочек является один (в случае если головные слова семантических цепочек совпадают) либо

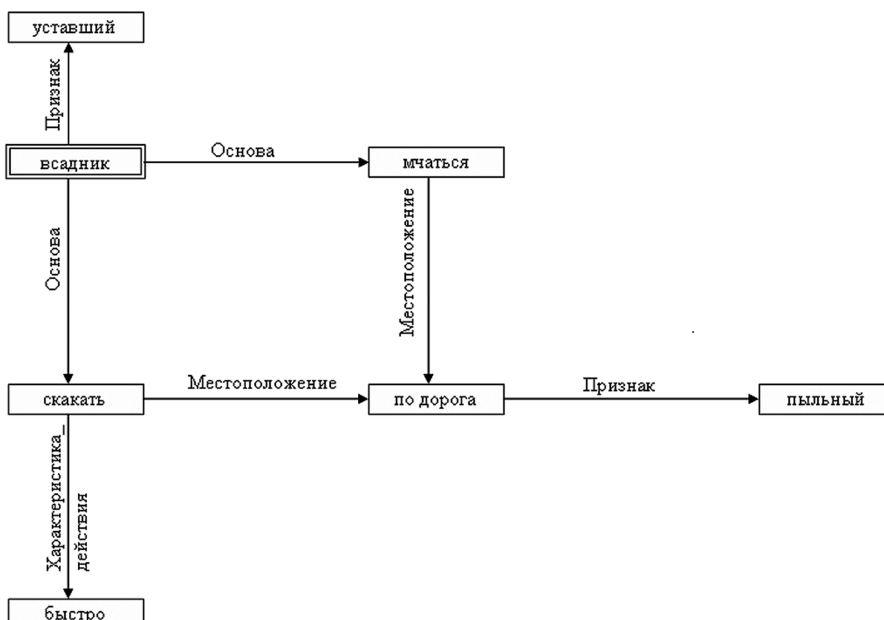


Рис. 2. Представление семантической цепочки в виде графа

два графа. Каждый граф, полученный в результате наложения семантических цепочек, является взвешенным и ориентированным. Узлы такого графа – это аргументы семантических связей, а ребра – семантические связи, связывающие эти аргументы. Каждое ребро такого графа имеет свой вес, определяющий количество вхождений конкретной семантической цепочки в анализируемый текст.

В результате наложения семантической цепочки $R_{\alpha\beta}^1 = \{R_1^1(\alpha_1^1, \beta_1^1), R_2^1(\alpha_2^1, \beta_2^1), \dots, R_n^1(\alpha_n^1, \beta_n^1)\}$, головным словом которой является α_1^1 , на семантическую цепочку $R_{\alpha\beta}^2 = \{R_1^2(\alpha_1^2, \beta_1^2), R_2^2(\alpha_2^2, \beta_2^2), \dots, R_m^2(\alpha_m^2, \beta_m^2)\}$ с головным словом α_1^2 ($m > n$) должна быть получена либо одна семантическая цепочка $R_{\alpha\beta}^* = \{R_1^*(\alpha_1^*, \beta_1^*), R_2^*(\alpha_2^*, \beta_2^*), \dots, R_k^*(\alpha_k^*, \beta_k^*)\}$ (в случае когда семантические зависимости, содержащие головные слова цепочек $R_{\alpha\beta}^1$ и $R_{\alpha\beta}^2$, совпадают, то есть $R_{\alpha\beta}^1(\alpha_1^1, \beta_1^1) = R_{\alpha\beta}^2(\alpha_1^2, \beta_1^2)$), либо две исходные семантические цепочки $R_{\alpha\beta}^1$ и $R_{\alpha\beta}^2$.

Семантическая цепочка $R_{\alpha\beta}^*$ (для случая когда $R_{\alpha\beta}^1(\alpha_1^1, \beta_1^1) = R_{\alpha\beta}^2(\alpha_1^2, \beta_1^2)$) формируется из цепочек $R_{\alpha\beta}^1$ и $R_{\alpha\beta}^2$ путем объединения соответствующих им графов: $R_{\alpha\beta}^* := R_{\alpha\beta}^1 \cup R_{\alpha\beta}^2$, при этом веса этих графов, представленные матрицами $W(R_{\alpha\beta}^1)$ и $W(R_{\alpha\beta}^2)$ соответственно, складываются: $W(R_{\alpha\beta}^*) := W(R_{\alpha\beta}^1) + W(R_{\alpha\beta}^2)$.

На рис. 3 представлена блок-схема алгоритма наложения семантических цепочек $R_{\alpha\beta}^1$ и $R_{\alpha\beta}^2$.

На рис. 4 показан пример наложения двух семантических цепочек $R_{\alpha\beta}^1$ (рис. 4а) и $R_{\alpha\beta}^2$ (рис. 4б), представленных в виде графов, у которых одинаковые узлы, являющиеся семантическими отношениями, обозначены одинаковыми литерами, а узел, являющийся семантическим отношением, содержащим головное слово цепочки, обведен в двойную рамку. На рис. 4с показана семантическая цепочка, полученная в результате

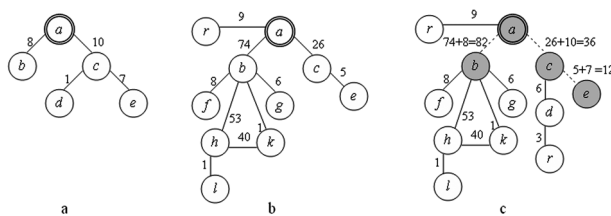


Рис. 4. Наложение семантических цепочек, представленных в виде графов

наложения цепочек $R_{\alpha\beta}^1$ и $R_{\alpha\beta}^2$. Вершины графа, являющиеся общими для графов, представляющих $R_{\alpha\beta}^1$ и $R_{\alpha\beta}^2$, выделены серым, а дуги, веса которых получились в результате суммирования соответствующих весов дуг $R_{\alpha\beta}^1$ и $R_{\alpha\beta}^2$, нарисованы пунктиром.

Применение алгоритма наложения семантических цепочек для автоматизации пополнения терминосистем

Предлагается создавать терминологическую систему, представляя ее в виде ориентированного, взвешенного графа, полученного в результате наложения всех семантических цепочек, построенных по анализируемому тексту семантическим анализатором. Такое представление терминосистемы позволит определять семантические связи между всеми терминами, содержащимися в анализируемых текстах. Терминологическую систему, построенную вышеописанным образом, можно постоянно пополнять путем анализа новых текстов. Это позволит хранить в системе новейшую информацию о терминах и связях между ними.

Значения веса ребер семантического графа было решено нормировать на диапазон от 0 до 1.

На основании анализа весов ребер графа, входящих в общее представление терминосистемы и образующих путь от одной вершины-термина до другой, можно делать предположение о корректности и достоверности построенной между терминами связи. Логично предположить, что большее значение весов обуславливает большую корректность и достоверность связи, определенной между терминами. Малое же значение весов может быть поводом для рассмотрения найденных связей между двумя терминами специалистом вручную.

Так как ребра графа, представляющего терминосистему, имеют веса и в качестве фактов рассматриваются семантические цепочки, то каждому факту в терминосистеме можно в соответствие поставить значение «веса факта», который, например, может определяться минимальным весом из всех ребер цепочки. Тогда для каждого факта в терминологической системе можно предположить его истинность и достоверность: очевидно, что факт с большим весом будет иметь большую достоверность с большей вероятностью.

Вышеописанный способ построения терминосистемы был программно реализован на языке

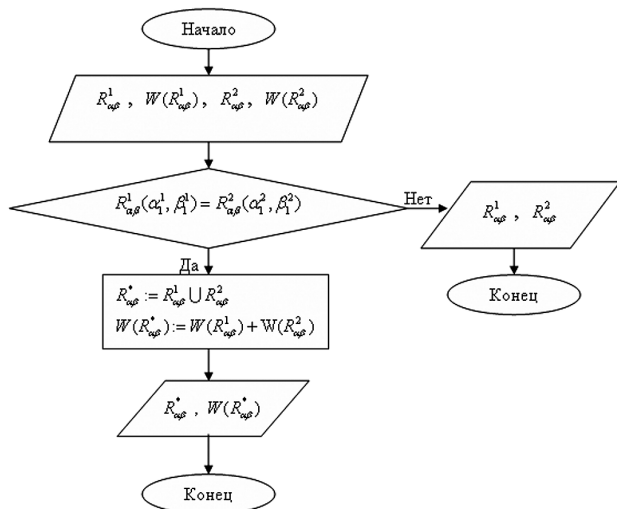


Рис. 3. Блок-схема алгоритма наложения семантических цепочек

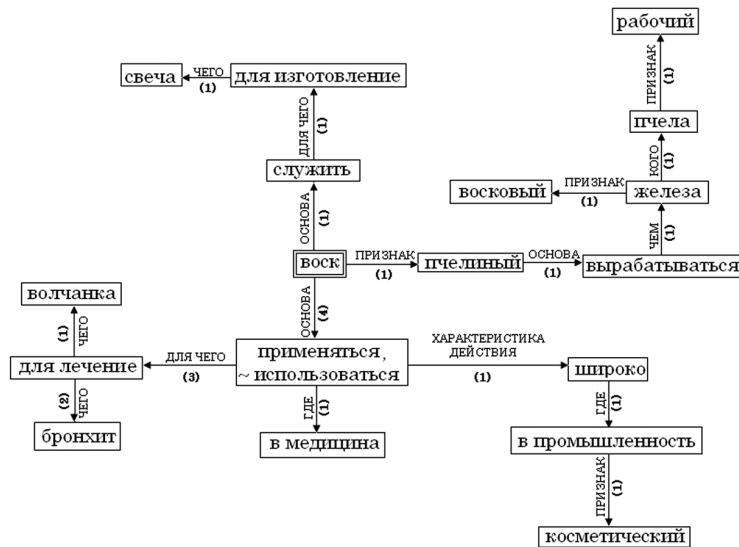


Рис. 5. Граф представления терминосистемы

программирования Java и в качестве составного модуля внедрен в вопросно-ответную систему, описанную в работе [11]. В качестве семантического анализатора, с помощью которого находятся семантические отношения и из которых впоследствии составляются семантические цепочки, использовалась программа-анализатор, алгоритм работы которой описан в статье [10].

Пример 3: Рассмотрим в качестве анализируемого текста следующие предложения, описывающие некоторые свойства термина «воск»: «Пчелиный воск вырабатывается восковыми железами рабочих пчел. Для лечения бронхитов используется воск. Воск широко применяется в косметической промышленности. Воск использовали в медицине. Воск служит для изготовления свечей. Воск применяется для лечения волчанки. Воск используется для лечения бронхитов».

Для этого текста по правилам, описанным в разделе «Алгоритм наложения семантических цепочек», будет создан ориентированный, взвешенный граф, представленный на рис. 5 (головное слово, являющееся термином, заносимым в систему, обведено двойной рамкой).

ЗАКЛЮЧЕНИЕ

В работе предложен способ создания и пополнения терминологической системы с помощью

разработанного и описанного алгоритма наложения семантических цепочек. Семантические цепочки специальным образом составляются из связанных семантических зависимостей, которые семантический анализатор строит для каждого анализируемого предложения. Связность семантических зависимостей предлагается определять условиями (1)–(4).

В отличие от наиболее распространенного в настоящее время способа пополнения терминосистем, основанного на использовании толковых и энциклопедических словарей, предлагаемый способ позволяет дополнять такие системы данными, содержащимися в текстовых источниках, на которые не накладываются дополнительные ограничения (помимо корректности описываемой информации и правильности составления предложений).

Предлагаемый в работе способ автоматического составления и пополнения терминосистемы программно реализован на языке программирования Java. Созданная терминосистема используется в алгоритме работы вопросно-ответной системы, описанной в статье [11].

В перспективе планируется использовать алгоритм построения и пополнения терминосистемы для автоматизации создания онтологии, которая будет интегрирована с вопросно-ответной системой [11].

* Работа выполнена при финансовой поддержке РГНФ в рамках научного проекта № 15–04–12029 «Программная разработка электронного ресурса с онлайн-версией русскоязычной вопросно-ответной системы».

СПИСОК ЛИТЕРАТУРЫ

- Арзамасцева И. В. Модели, методы и средства разработки лингвистического обеспечения проектных репозитивов САПР: Дисс. ... канд. техн. наук. Ульяновск, 2011. 270 с.
- Гринев С. В. Введение в терминоведение. М.: Московский Лицей, 1993. 309 с.
- Мюллер Ю. Э. Применение онтологий для создания терминосистем (на примере терминологии логистики) // II Congreso Internacional «La lengua y literatura rusas en el espacio educativo internacional: estado actual y perspectivas». Т. II. Мадрид: Rubinos-1860, S. A, 2010. С. 2038–2042.

4. Ермаков А. Е. Автоматизация онтологического инжиниринга в системах извлечения знаний из текста // Труды международной конференции «Диалог 2008» [Электронный ресурс]. Режим доступа: <http://www.dialog-21.ru/digests/dialog2008/materials/html/23.htm>
5. Ефремова Т. Ф. Новый словарь русского языка (толково-словообразовательный). М.: Дрофа: Русский язык, 2000. 1233 с.
6. Леонова С. А. Терминосистема «автоматизированный электропривод» в английском и русском языках: синхронно-диахронный анализ: Дисс. ... канд. филол. наук. М., 2012. 257 с.
7. Майтлова А. В. Терминосистема предметно-специального языка «Банковское дело» в лингвокогнитивном аспекте – на материале русского и немецкого языков: Дисс. ... канд. филол. наук. М., 2008. 257 с.
8. Минаков И. А. Системный анализ, онтологический синтез и инструментальные средства обработки информации в процессах интеграции профессиональных знаний: Автореф. дисс. ... д-ра техн. наук. Самара, 2007. 43 с.
9. Митрофанова О. А., Константинова Н. С. Онтологии как системы хранения знаний // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы». 2008. 54 с.
10. Мочалова А. В. Алгоритм семантического анализа текста, основанный на базовых семантических шаблонах с удалением // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 5. С. 126–132.
11. Мочалова А. В., Мочалов В. А. Интеллектуальная вопросно-ответная система // Информационные технологии. 2011. № 5. С. 6–12.
12. Найханова Л. В. Методы и модели автоматического построения онтологий на основе генетического и автоматного программирования: Дисс. ... д-ра техн. наук. М., 2009. 451 с.
13. Нечипоренко А., Русин А. Система автоматизированного извлечения знаний из текстов на естественном языке // Труды международной научно-технической конференции «Информационные системы и технологии – 2003». Новосибирск, НГТУ, 2003 [Электронный ресурс]. Режим доступа: <http://www.noolab.ru/index.php?id=stat&show=18>
14. Пекар В. И. Автоматическое пополнение специализированного тезауруса // Труды международной конференции «Диалог 2002» [Электронный ресурс]. Режим доступа: <http://www.dialog-21.ru/en/digest/archive/2002/?year=2002&vol=22725&id=7631>
15. Рубашкин В. Ш. Онтологическая семантика. Знания. Онтологии. Онтологически ориентированные методы информационного анализа текстов. М.: ФИЗМАТЛИТ, 2013. 348 с.
16. Сокирко А. В. Семантические словари в автоматической обработке текста: По материалам системы ДИАЛИНГ: Дисс. ... канд. техн. наук. М., 2001. 120 с.
17. Agramaki E., Imai T., Kashiwagi M., Kajino M., Miyo K. and Ohe K. Toward medical ontology using Natural Language Processing. Available at: <http://www.m.u-tokyo.ac.jp/medinfo/ont/paper/2005-ara.-1.pdf>
18. Novy E., Knight K., Junk M. Large Resources. Ontologies (SENSUS) and Lexicons. Available at: www.isi.edu/natural-language/projects/ONTOLOGIES.html
19. Morris J., Hirst G. Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of the Text // Computational Linguistics, 1991. № 17 (1). P. 21–45.
20. Roussopoulos N. D. A semantic network model of data bases. – TR No 104, Department of Computer Science, University of Toronto, 1976.

Mochalova A. V., Petrozavodsk State University (Petrozavodsk, Russian Federation)

CREATION AND UPDATE OF TERMINOLOGICAL SYSTEMS BY MEANS OF SEMANTIC ANALYZER

An automated creation and update of terminological systems is an urgent challenge today. Such systems are used to solve a wide range of problems concerning the process of text analysis. The problem associated with these systems is substantiated by the need to permanently update current information. In the paper we propose a method, which automatically creates and updates terminological systems with the help of an algorithm. The algorithm overlaps semantic chains that are derived from a set of semantic dependencies built by a semantic analyzer applied to every sentence of the text. The advantage of this method, developed to create and update terminological systems, is based on the fact that any Russian authentic textual data may be used as an analyzed text. Encyclopedia dictionaries or glossaries can also be studied with the help of the analyzer. This method is instrumental in the update of the system with the information from new textual sources: scientific papers, books, scientific reports, news or analytical reviews.

Key words: terminological systems, update of terminological systems, semantic chains, semantical analyzer

REFERENCES

1. Arzamastseva I. V. *Modeli, metody i sredstva razrabotki lingvisticheskogo obespecheniya proektnykh repozitoriev SAPR: Diss. ... kand. tekhn. nauk* [Models, methods and tools for development of linguistic support of project repositories of CAD. PhD. tech. sci. diss.]. Ulyanovsk, 2011. 270 p.
2. Grinev S. V. *Vvedenie v terminovedenie* [Introduction to terminology studies]. Moscow, Lyceum of Moscow Publ., 1993. 309 p.
3. Myuller Yu. E. Using ontologies for creating term systems (on an example of logistics terminology) [Primenenie ontologii dlya sozdaniya terminosistem (na primere terminologii logistiki)]. *II Congreso Internacional "La lengua y literatura rusas en el espacio educativo internacional: estado actual y perspectivas"*. Vol. II. Madrid: Rubinos-1860, S. A, 2010. P. 2038–2042.
4. Ermakov A. E. Automatization of an ontological engineering for systems of knowledge mining in text [Avtomatizatsiya ontologicheskogo inzhiniringa v sistemakh izvlecheniya znaniy iz teksta]. *Trudy mezhdunarodnoy konferentsii "Dialog 2008"*. Bekasovo, 2008. Available at: <http://www.dialog-21.ru/digests/dialog2008/materials/html/23.htm>
5. Efremova T. F. *Novyy slovar' russkogo yazyka (tolkovo-slovoobrazovatel'nyy)* [New dictionary of the Russian language (explanatory-word formative)]. Moscow, Drofa, The Russian Language Publ., 2000. 1233 p.
6. Leonova S. A. *Terminosistema "Avtomatizirovannyy elektropriwod" v angliyskom i russkom yazykakh: sinkhronno-diakhronnyy analiz: Dis. ... kand. filol. nauk* [Term system "Automatic electric drive" in English and Russian languages: synchronous-diachronous analysis. PhD. phil. sci. diss.]. Moscow, 2012. 257 p.

7. М а y т о в а А. В. *Terminosistema predmetno-spetsial'nogo yazyka "Bankovskoe delo" v lingvokognitivnom aspekte – na materiale russkogo i nemetskogo yazykov: Diss. ... kand. filol. nauk* [Term system of the subject-specific language "Banking business" in a linguo-cognitive aspect – as exemplified in the Russian and German languages. PhD. phil. sci. diss.]. Moscow, 2008. 257 p.
8. М и n а k о в I. A. *Sistemnyy analiz, ontologicheskii sintez i instrumental'nye sredstva obrabotki informatsii v protsessakh integratsii professional'nykh znaniy: Avtoref. diss. ... d-ra tekhn. nauk* [System analysis, ontological synthesis and data processing tools in the processes of professional knowledge integration. Dr. tech. sci. diss.]. Samara, 2007. 43 p.
9. М и t r o f a n o v a O. A., К o n s t a n t i n o v a N. S. Ontologies as systems of data storage [Ontologii kak sistemy khraneniya znaniy]. *Vserossiyskiy konkursnyy otbor obzorno-analiticheskikh statey po prioritetnomu napravleniyu "Informatsionno-telekommunikatsionnye sistemy"* [All-Russian competitive selection of the review analytical papers on the priority direction "Information telecommunication systems"], 2008.
10. М o c h a l o v a A. V. Algorithm of semantical text analysis based on the basical semantical templates with removal [Algoritm semanticheskogo analiza teksta, osnovanny na bazovykh semanticheskikh shablonakh s udaleniem]. *Nauchno-tekhnicheskiiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and Technical Journal of Information Technologies, Mechanics and Optics], 2014. № 5. P. 126–132.
11. М o c h a l o v a A. V., М o c h a l o v V. A. Intellectual question-and-answer system [Intellektual'naya voprosno-otvetnaya sistema]. *Informatsionnye tekhnologii* [Information technologies], 2011. № 5. P. 6–12.
12. N a y k h a n o v a L. V. *Metody i modeli avtomaticheskogo postroyeniya ontologiy na osnove geneticheskogo i avtomatnogo programmirovaniya: Diss. ... d-ra tekhn. nauk* [Methods and models for automated ontologies construction basing on genetic and automata-based programming. Dr. tech. sci. diss.]. Moscow, 2009. 451 p.
13. N e c h i p o r e n k o A., R u s i n A. A system for automated knowledge extraction from natural-language texts [Sistema avtomatizirovannogo izvlecheniya znaniy iz tekstov na estestvennom yazyke]. *Trudy mezhdunarodnoy nauchno-tekhnicheskoy konferentsii "Informatsionnye sistemy i tekhnologii – 2003"*. Novosibirsk, NSTU Publ., 2003. Available at: <http://www.noolab.ru/index.php?id=stat&show=18>
14. P e k a r V. I. Automated updating of the problem-specific thesaurus [Avtomaticheskoe popolnenie spetsializirovannogo tezaurus]. *Trudy mezhdunarodnoy konferentsii "Dialog 2002"*. Available at: <http://www.dialog-21.ru/en/digest/archive/2002/?year=2002&vol=22725&id=7631>
15. R u b a s h k i n V. S h. *Ontologicheskaya semantika. Znaniya. Ontologii. Ontologicheski orientirovannyye metody informatzionnogo analiza tekstov* [Ontological semantics. Knowledge. Ontologies. Ontologically-oriented methods of information text analysis]. Moscow, FIZMATLIT Publ., 2013. 348 p.
16. S o k i r k o A. V. Semanticheskie slovari v avtomaticheskoy obrabotke teksta: Po materialam sistemy DIALING. Diss. ... kand. tekhn. nauk [Semantical dictionaries in automated text processing: as exemplified by the DIALING system. PhD. tech. sci. diss.]. Moscow, 2001. 120 p.
17. A r a m a k i E., I m a i T., K a s h i w a g i M., K a j i n o M., M i y o K. and O h e K. Toward medical ontology using Natural Language Processing. Available at: <http://www.m.u-tokyo.ac.jp/medinfo/ont/paper/2005-ara..-1.pdf>
18. H o v y E., K n i g h t K., J u n k M. Large Resources. Ontologies (SENSUS) and Lexicons. Available at: www.isi.edu/natural-language/projects/ONTOLOGIES.html
19. M o r r i s J., H i r s t G. Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of the Text // *Computational Linguistics*. 1991. № 17 (1). P. 21–45.
20. R o u s s o p o u l o s N. D. A semantic network model of data bases. – TR No 104, Department of Computer Science, University of Toronto, 1976.

Поступила в редакцию 21.01.2015