



УДК 004.822:514

КОНСТРУИРОВАНИЕ ОТНОШЕНИЯ ПСЕВДОХРОНОЛОГИЧЕСКОГО ПОРЯДКА В ТЕРМИНОЛОГИЧЕСКИХ СЕТЯХ

Мальковский М.Г., Соловьев С.Ю.

*Факультет вычислительной математики и кибернетики,
Московский государственный университет имени М.В. Ломоносова,
г. Москва, Россия*

malk@cs.msu.su

soloviev@glossary.ru

Рассматривается жизненный цикл терминов научной и деловой лексики. Вводится понятие хронологической шкалы терминов. Рассматривается задача восстановления хронологических связей на заданном структурированном множестве терминов. Предлагается для восстановления использовать приемы научного познания, которые с одной стороны используются для конструирования терминосистем, а с другой стороны позволяют зафиксировать порядок появления терминов в научно-деловом обороте. Приводятся образцы продукционных правил, ориентированных на выявление хронологических связей. Предлагается метод конструирования хронологической шкалы по результатам функционирования продукционной системы.

Ключевые слова: термин, терминологическая сеть, хронология, продукционная система.

Введение

В научном и деловом обращении термины, хотя и существуют относительно самостоятельно, главным образом используются для именованя понятий. Известно, что с течением времени¹ термины:

- вводятся в обращение (появляются);
- замещаются синонимами;
- изменяют смысл;
- выбывают из обращения (устаревают).

Так в 1990-е годы появились новое понятие и новый термин “Евро”. Примером типичного замещения является вытеснение из научно-технических текстов термина “ЭВМ” в пользу термина “Компьютер”. Если под изменением смысла понимать изменение объема понятия, то термин “Еврозона” изменялся несколько раз. За последние 20 лет в категорию устаревших перешли термины “Ленинград”, “Единый социальный налог” и “Евродоллар”, а также устарела целая терминосистема “Научный коммунизм”. Заметим, что со временем устаревшие термины (если не принять меры) могут либо породить проблему неоднозначного понимания накопленных электронных текстов, либо

загромождать веб-пространство, в котором отсутствует хронология документов.

Разработка глобальных терминологических баз знаний существенно затрудняется необходимостью моделировать хронологические связи и временные характеристики терминов. В последующих рассуждениях мы исходим из того, что каждый термин однажды изобретается человеком и, после определенного периода “борьбы за признание”, включается в активный научно-деловой оборот. В некоторых случаях возникновение термина удается датировать весьма точно. Так, термин “Наркоманическая зависимость” как “зависимость больного от присутствия или отсутствия в его организме алкоголя” введен ВОЗ в 1964 году [Дьяченко, 2002]. В других случаях, как это произошло, скажем, с “Ноосферой” В.И.Вернадского, история зарождения термина растягивается на десятилетия.

Настоящая работа посвящена задаче восстановления по косвенным признакам хронологии появления терминов, вопросы изменения смысла и устаревания терминов практически не рассматриваются.

1. Терминологические сети

Терминологическая сеть [Мальковский и др., 2012] есть продукт деятельности коллектива

¹ Жизненный цикла термина в пределах одной терминосистемы называется терминогенезом [Кокорина, 2011].

научных редакторов, занимающихся систематизацией терминов и их определений. С формальной точки зрения терминологическая сеть представляет собой ориентированный граф, узлы которого соответствуют терминам², а дуги – экземплярам бинарных отношений из заранее фиксированного набора допустимых отношений. В терминологических сетях:

- каждый узел A можно рассматривать как четверку $\langle N, T, S, D \rangle$, где

$N = N(A)$ – необязательное понятийное имя,

$D = D(A)$ – определение термина,

$T = T(A)$ – выделенный термин из множества синонимов $S = S(A)$.

- каждая дуга $(A, B)_R$ есть упорядоченная пара узлов A и B , помеченная символом отношения R ; если для дуги не оговаривается тип отношения, то в записи такой дуги метка опускается;

- набор допустимых отношений обязательно содержит родовидовое отношение P , которому соответствуют дуги $(A, B)_P$, где A – вид, B – род;

- понятийным узлом называется узел, в который заходит хотя бы одна дуга;

- потомками понятийного узла B называются узлы A , связанные с B дугами (A, B) ;

- предками понятийного узла A называются узлы B , связанные с узлом A дугами (A, B) ;

- каждый понятийный узел A имеет уникальное имя $N(A)$, которое также служит наименованием понятия; как правило, наименование понятия есть общее наименование составляющих объектов.

Выявление терминов, пригодных для образования понятийных узлов, а также установление отношений между терминами составляет основу интеллектуальной деятельности научных редакторов. В своей работе редакторы используют определения терминов, справочную литературу и собственные знания.

При анализе терминологических сетей существенно используются специальные отношения между терминами, узлами и дугами. Приведем эти отношения.

Будем говорить, что (многословный) термин x подчинен (многословному) термину y , если термин x является развитием термина y . Примерами отношения подчиненности являются следующие пары терминов:

$x =$ “Промышленные аварии” и $y =$ “Аварии”;

$x =$ “Централизованная библиотечная система” и

$y =$ “Библиотечные системы”;

$x =$ “Скорость света в вакууме” и $y =$ “Вакуум”.

Отношение подчиненности для пары терминов, связанных дугой, отвечает синтаксическому способу терминоподчинения [Гринев-Гриневиц, 2008]. С формальной точки зрения термин x , состоящий из слов x_1, \dots, x_g , подчинен термину y , состоящему из

слов y_1, \dots, y_h , если существует однозначная функция

$$f : \{y_1, \dots, y_h\} \rightarrow \{x_1, \dots, x_g\} \quad (1)$$

такая, что для всех $i = 1, \dots, h$ слова y_i и $f(y_i)$ отличаются формальными суффиксами.

2. Хронологическая шкала

Будем полагать существование порядковой хронологической шкалы терминов, позволяющей судить о порядке их появления. Строго говоря, шкала моделируется симметричным и транзитивным бинарным отношением предшествования на множестве узлов терминологической сети: если A и B два узла, то запись $A \leq B$ означает, что термин A появился не позднее термина B . Соответственно отношение $A = B$ определяется как $(A \leq B) \ \& \ (B \leq A)$.

Понимая под отношением множество упорядоченных пар, будем называть хронологической связью одну пару из отношения. Примерами хронологических связей между терминами являются:

“Здание” \leq “Кровля”,

“Алкоголь” \leq “Наркоманическая зависимость”,

“Облигация” \leq “Дисконтная облигация”,

“Дисконтная облигация” \leq

\leq “Облигация с глубоким дисконтом”,

“Облигация” \leq “Облигация с глубоким дисконтом”.

Хронологическая шкала терминов открывает новые возможности для конструирования, анализа и моделирования эволюционных процессов в терминосистемах.

3. Подход к восстановлению хронологических связей

По сути дела терминологическая сеть представляет собой частично формализованную структуру, в которой бинарные связи между терминами фиксируются дугами семантической сети, а остальная информация содержится в определениях терминов (в словарных статьях). Кроме того, принятый способ организации данных позволяет относить к формальной части структуры синонимы и аббревиатуры терминов.

Терминологическая сеть, будучи существенно статической структурой данных, тем не менее позволяет с определенной достоверностью установить некоторые связи на хронологической шкале терминов. Основной подход к реконструкции хронологических связей состоит в том, что при конструировании [новых] понятий и соответствующих им терминов существенно используются общенаучные методы анализа и абстрагирования, а также приемы сравнения и аналогии [Степин, 2003]. Применительно к понятиям упомянутые методы и приемы имеют особенности, позволяющие делать выводы о хронологических связях.

² Термины в свою очередь являются наименованиями понятий.

Анализ понятий – метод конструирования новых понятий в результате разложения известного понятия (α -понятия) на его составляющие. В терминологической сети новым подпонятиям соответствуют потомки α -узла. С точки зрения хронологических связей новые понятия появляются позже α -понятия.

Абстрагирование – метод конструирования новых понятий (β -понятий) посредством отвлечения от несущественных деталей, присущих известным понятиям (γ -понятиям). В терминологической сети β -понятиям соответствуют предки γ -узлов. С точки зрения хронологических связей β -понятия появляются позже γ -понятий.

Сравнение понятий – прием выявления общих признаков заданных понятий.

Вывод по аналогии – прием выявления новых признаков понятий, при котором на основе сходства понятий по одним признакам заключается их сходство по другим признакам.

4. Правила восстановления хронологических связей

В качестве исходных данных в задаче восстановления хронологических связей используется связный фрагмент терминологической сети отвечающей некоторой терминсистеме. Для выявления отдельных хронологических связей предлагается использовать продукционные правила вывода, некоторые из которых приводятся далее.

Правило 1. ЕСЛИ (а) $(A, B)_P$ – дуга, И (б) множества $S(A)$ и $S(B)$ содержат только термины $T(A)$ и $T(B)$, И (в) $T(A)$ подчинен $T(B)$, ТО $B \leq A$.

Правило 2. ЕСЛИ (а) B – узел, И (б) для всех A_i (таких, что $(A_i, B)_P$ дуга терминологической сети) множество $S(A_i)$ содержит термин s_i подчиненный термину $T(B)$, ТО $B \leq A_i$.

Правило 3. ЕСЛИ (а) $(A, B)_P$, $(B', B)_X$, $(A', A)_Y$, $(A', B')_P$ – дуги, причем $X = Y \neq P$, И (б) множества $S(A)$ и $S(B)$ содержат только термины $T(A)$ и $T(B)$, И (в) $T(A)$ не подчинен $T(B)$, ТО $A \leq B$ (см. рис.1 левая часть).

Правило 4. ЕСЛИ (а) $(A, B)_P$ – дуга, И (б) множества $S(A)$ и $S(B)$ содержат только термины $T(A)$ и $T(B)$, И (в) $N(A)$ не подчинен $N(B)$, ТО $A \leq B$.

Правило 5. ЕСЛИ (а) $(A, B)_P$ – дуга, И (б) существует цепочка дуг $(B, B^{(1)})_P, \dots, (B^{(n-1)}, B^{(n)})_P$, И (в) $T(B^{(n)}) = \text{“Идеализированный объект”}$, ТО $A \leq B$.

Правило 6. ЕСЛИ (а) (A, B) – дуга, И (б) для любой пары терминов (s_A, s_B) из $S(A) \times S(B)$ термин s_A не подчинен s_B , ТО $A \leq B$. (рис.1 правая часть).

Правило 7. ЕСЛИ (а) $(A, B)_X$, $(A', B)_X$ – дуги, с одинаковыми метками, И (б) в сети отсутствуют иные дуги, исходящие из A и A' , И (в) $A \leq B$, ТО $A' \leq B$.

Правило 8. ЕСЛИ (а) $(A, B)_X$, $(A', B)_X$ – дуги с одинаковыми метками, И (б) в сети отсутствуют иные дуги, исходящие из A и A' , И (в) $B \leq A$, ТО $B \leq A'$.

Замечание 1. Приведенные правила являются примерами продукций, ориентированных на формирование хронологической шкалы терминов для выбранного фрагмента терминологической сети. Естественно, правила такого рода должны иметь числовые коэффициенты уверенности (CF), значения которых в конечном итоге зависят от выбранной схемы правдоподобных рассуждений [Моросанова и др., 2012], а на выбранном уровне описания можно зафиксировать лишь общие свойства этих коэффициентов:

- правила 1-8 не являются абсолютно достоверными,
- CF правила 3 больше CF правила 1 и т.д.

Результатом функционирования продукционной системы является матрица коэффициентов уверенности фактов вида $A \leq B$. Матрица попарных оценок открывает возможность использовать для построения искомой порядковой шкалы алгоритмы многомерного шкалирования [Терехина, 1986]. По построению в матрице содержатся незаполненные позиции, количество которых можно сократить посредством транзитивного замыкания:

ЕСЛИ $A_1 \leq A_2$ [CF=a] И $A_2 \leq A_3$ [CF=b],
ТО $A_1 \leq A_3$ [CF=comb(a,b)],

где *comb* – операция комбинирования коэффициентов уверенности.

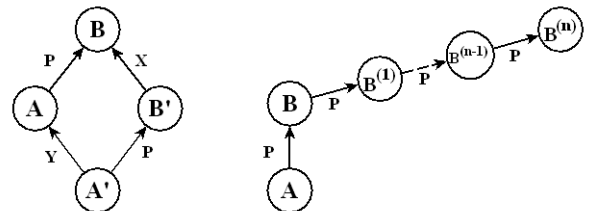


Рисунок 1 – Условия правил 3 и 5

Замечание 2. Правила 1 и 2 исходят из того, что понятия A и A_i появились в результате анализа ранее существовавшего α -понятия B . Предполагается, что анализ состоит в выделении подпонятий и связанных с ними сервисными понятиями.

Замечание 3. Правила 3-6 исходят из того, что β -понятие получено в результате абстрагирования как новое знание, обобщающее ранее известные γ -понятия A и B' . Особое место занимает правило 5, в котором примененный метод познания идентифицируется смыслом узла $B^{(n)}$.

Замечание 4. Правила 7 и 8 соответствуют выводам по аналогии, в которых хронологическая связь, полученная тем или иным способом для узлов дуги $(A, B)_X$, распространяется на узлы дуги $(A', B)_X$.

Замечание 5. Некоторые дуги вида $(A, B)_P$, удовлетворяют условиям правил 4-6. При этом в

сети возможна ситуация, зафиксированная в следующем правиле:

Правило 9. ЕСЛИ $(A,B)_P$ И $(A',B)_P$ – различные дуги, удовлетворяющие условиям правил 4-6, ТО $A \leq B$ и $A' \leq B$.

Условие правила 9 фиксирует прием сравнения понятий A и A' . Особенность правила 9 состоит в том, что его коэффициент уверенности выше коэффициентов каждого из правил 4-6.

Замечание 6. Поскольку введенное отношение хронологического предшествования является симметричным, то для любого фрагмента терминологической сети искомая хронологическая шкала может быть построена как шкала одновременно возникших терминов (терминов, связанных отношением $=$). Естественно, такое решение не является информативным, однако оно открывает подход к построению нетривиальных хронологических шкал дивизимным методом [Жамбю, 1988] разделения множества терминов на хронологически упорядоченные и более мелкие составляющие.

Замечание 7. Некоторые термины (и связанные с ними понятия) имеют точные дату возникновения. Такие термины представляют особую ценность, поскольку их можно использовать для отбраковки неверно построенных хронологических связей.

Заключение

1. Описанный подход не влияет на топологию терминологической сети: все узлы и дуги сети остаются без изменений, что ограничивает область применения описанной методологии. Моделирование процессов замещения, изменения и устаревания терминов требует дополнительных исследований. Представляется, что в первую очередь следует определиться с понятийным аппаратом терминогенеза.

2. Предложенный подход в лучшем случае позволяет конструировать псевдохронологические связи, лишь отчасти совпадающие с истинными связями. Исчерпывающий подход к решению задачи конструирования хронологической шкалы состоит в поддержке специальной структуры, описывающую эволюцию каждого конкретного термина. Однако это невозможно по двум причинам: (а) задача нуждается в точной и плодотворной постановке, и (б) задача является слишком трудоемкой.

Библиографический список

- [Гринев-Гриневиц, 2008] Введение в терминографию / С.В.Гринев-Гриневиц. – М.: ЛИБРОКОМ, 2009.
[Дьяченко, 2002] Словарь авторских терминов, понятий и названий в медицине / А.П.Дьяченко. – М.: Триада-Х, 2002.
[Жамбю, 1988] Иерархический кластер-анализ и соответствия / М.Жамбю. – М.: Финансы и статистика, 1988.
[Мальковский и др., 2012] Мальковский М.Г., Терминологические сети / М.Г.Мальковский, С.Ю.Соловьев // Материалы конференции OSTIS-2012. – Мн.: БГУИР, 2012. С. 77-82

[Моросанова и др., 2012] Моросанова Н.А., Формальные свойства схемы Шортлиффа / Н.А.Моросанова, С.Ю.Соловьев // Управление большими системами, выпуск 36. - М.: ИПУ РАН, 2012. С. 5-38

[Кокорина, 2011] Терминосистема концептуальной области «Археология» / Ю.Г.Кокорина. – М.: ЛИБРОКОМ, 2011.

[Степин, 2003] Теоретическое знание / В.С.Степин. – М.: Прогресс-Традиция, 2003.

[Терехина, 1986] Анализ данных методами многомерного шкалирования / А.Ю.Терехина. – М.: Наука, 1986.

THE DESIGNING OF PSEUDO CHRONOLOGICAL ORDER RELATION IN THE TERMINOLOGICAL NETWORKS

Malkovsky M.G., Soloviev S.Y.

Lomonosov MSU, Moscow, Russia

malk@cs.msu.su

soloviev@glossary.ru

In this article we

- consider we consider the terms life cycle;
- introduce the notion of term chronological scale;
- consider the problem of reconstructing the historical ties in a given structured sets of terms;
- propose to use for the restoration techniques of scientific knowledge, which on one hand is used for the construction of terminological, on the other hand allow you to fix the order of appearance in terms of scientific and business turnover;
- give examples of production rules, aimed at identifying historical ties;
- propose a method of constructing a chronological scale according to the functioning of a production system.

Introduction

The problem of the evolution of terms is seen as a difficult task, which has a subtask of designing a term chronological scale.

Main Part

Object of research are terminology network, which reflect the relationship between terms. The terms chronological scale must satisfy the axioms of symmetric, and transitive. The method of building chronological ties between the terms is based on the recognition methods of scientific knowledge, which were the basis of terminological structure. These methods are called: the method of analysis of concepts, the method of abstraction, the method of comparison and analogy method.

Methods of scientific knowledge generate production rules, some of which are presented in the article. The production system builds a matrix of comparative assessments of terminology ties. Justification production rules boil down to different variants of the method of scientific knowledge. Ultimately, the task of designing terminology scale is reduced to the hierarchical clustering.

Conclusion

We give a brief overview of the problem of modeling subtasks of terms evolution.