

УДК 528.23

И.В. Евглевский

ФОРМАЛИЗАЦИЯ ЗНАНИЙ ПРИ ПОМОЩИ ФУНКЦИОНАЛЬНЫХ КЛАССОВ ЕСТЕСТВЕННОГО ЯЗЫКА НА ПРИМЕРЕ ФОРМИРОВАНИЯ КООРДИНАТНОЙ СРЕДЫ ГЕОИНФОРМАЦИОННОЙ СИСТЕМЫ

Приведены основы формализации знаний на примере формирования координатной среды для географических информационных систем. Описаны функциональные классы естественного языка, на основе которых формализуются знания. На примерах отображения поверхностей показан процесс построения RX-кодов, синтагматических цепей и семантических сетей в целях моделирования предметной области для формирования координатной среды геоинформационных систем.

Введение

Одним из сложнейших вопросов построения информационных систем (в том числе геоинформационных) является вопрос организации процесса оперирования компьютером не только данными, но и понятийными знаниями (создание базы знаний). В информационных системах под знаниями понимают сложноорганизованные данные, содержащие фактографическую и семантическую информацию, имеющую в своем составе встроенные процедуры обработки [1]. В отличие от данных знаниям присущи следующие особенности [2]:

- внутренняя интерпретируемость;
- структурированность;
- связность;
- семантическая метрика;
- активность.

Автоматизация вывода знаний требует адекватных действий по их формализации (представлению): если машинные программы смогут аккумулировать и применять знания о конкретной предметной области, то они достигнут высочайшего уровня работы [3].

Геоинформационная система формирует ответ на запрос в процессе реализации понятийного аппарата предметной области, охватывающей закономерности отношений пространственно-временных данных, интегрированных в географическую информацию. Вывод ответа на запрос в геоинформационной системе осуществляется в рамках системы управления базой знаний, объединяющей правила формализации, хранения, обработки и вывода информации. Соответственно в системе управления базой знаний реализуются три основные группы моделей ее функционирования [4]:

- 1) модели представления знаний о предметной области (способы формализации знаний);
- 2) модели установления семантического соответствия (способы установления близости запроса и фрагмента знаний);
- 3) модели логического вывода (способы получения новых знаний на основе хранящейся в базе знаний информации).

Первая группа моделей требует творческого подхода к созданию, наиболее сложна в практической реализации и наименее изучена в плане применения ее в геоинформационных технологиях. Универсальным языком представления знаний в геоинформационных системах может выступать естественный язык как основа формализации знаний, например, на предикатном и реляционном языках (искусственных языках), нашедших широкое применение в системах управления базами знаний в геоинформационных системах.

Представление знаний при помощи функциональных классов естественного языка

Процедура описания знаний основана на инженерии знаний и включает в себя правила идентификации объектов предметной области и связей между ними, составляющих формаль-

ные предложения на некотором языке, элементарными объектами которого являются первичные признаки или понятия [5].

На примере процесса формирования координатной среды геоинформационной системы рассмотрим основные способы формализации знаний при помощи функциональных классов естественного языка.

Основные функциональные классы естественного языка

Естественные языки складываются в процессе коммуникативной деятельности людей и являются, прежде всего, средством общения и передачи информации [6].

К основным функциональным классам естественного языка относятся понятия, имена и отношения, достаточно полно раскрывающие семантическую основу любого речевого высказывания.

Одна из форм отражения мира на ступени познания, связанная с применением языка, форма (способ) обобщения предметов и явлений называется понятием [7]. Обобщенные по некоторым совокупным признакам понятия можно объединить в классы. Примерами геопространственных понятий-классов являются: «эллипсоид», «поверхность», «геодезическая линия» и т. п.

Выражение языка, обозначающее некоторый предмет и однозначно идентифицирующее его среди подобных предметов данного понятия-класса, называется именем [7]. Примерами имен служат «эллипсоид Красовского», «поверхность, ограниченная северной широтой 50°, южной широтой 30°, западной долготой 20°, восточной долготой 30°», «геодезическая линия 1-2» и т. п.

Установление связей на множестве понятий-классов и идентифицированных понятий производится при помощи отношений [7]. Например, связь между понятием «поверхность» и «геодезическая линия» реализуется посредством отношения «лежит».

Общий алгоритм вычисления геодезических проекций

Возможность установления математических закономерностей решения целого ряда геодезических задач (выбор геодезических проекций, введение оптимальных систем координат, перевычисление координат точек в различных системах плоских прямоугольных координат и т. п.) в рамках геоинформационной системы позволяет отнести формирование координатной среды к разряду монотонных задач, для которых результат может быть описан и классифицирован.

Процесс формирования координатной среды для геоинформационной системы подразумевает организацию информационного пространства, допускающего различную степень обобщения и вычленения составляющих его элементов, с вполне определенной и достаточной точностью, и предполагает применение математических методов его отображения [8]. Таким образом, выявление связей между понятиями информационного координатно-временного пространства и составляет суть моделирования представления знаний о предметной области формирования координатной среды.

Пусть на двух регулярных поверхностях внутри обеих отображаемых областей заданы изометрические координаты начальной точки, определенные как начальные параметры отображения $u_{01}, v_{01}, u_{02}, v_{02}$. Тогда координаты любой точки внутри односвязных областей получают значения

$$\left. \begin{aligned} u_1 &= u_{01} + \Delta u_1; v_1 = v_{01} + \Delta v_1; \\ u_2 &= u_{02} + \Delta u_2; v_2 = v_{02} + \Delta v_2. \end{aligned} \right\} \quad (1)$$

Аналитические функции прямого и обратного конформного отображения односвязных поверхностей можно записать в виде

$$\left. \begin{aligned} u_2 + iv_2 &= f[u_{01} + \Delta u_1; i(v_{01} + \Delta v_1)]; \\ u_1 + iv_1 &= \varphi[u_{02} + \Delta u_2; i(v_{02} + \Delta v_2)]. \end{aligned} \right\} \quad (2)$$

Пусть функции (2) и размеры односвязных областей допускают разложение в бесконечно убывающий ряд по формуле Тейлора правых частей уравнений (2), тогда для координат (1) получаем общие в любой конформной проекции уравнения связи [8]:

$$\left. \begin{aligned} \Delta u_2 &= u_2 - u_{02} = \sum_{j=1}^n C_j P_j, \\ \Delta v_2 &= v_2 - v_{02} = \sum_{j=1}^n C_j Q_j. \end{aligned} \right\} \quad (3)$$

$$\left. \begin{aligned} \Delta u_1 &= u_1 - u_{01} = \sum_{j=1}^n C_j' P_j', \\ \Delta v_1 &= v_1 - v_{01} = \sum_{j=1}^n C_j' Q_j'. \end{aligned} \right\} \quad (4)$$

где C_j – коэффициенты разложений, полученные как частные производные от функции f по координатам u_1, v_1 , а C_j' – коэффициенты разложений, полученные как частные производные от функции φ по координатам u_2, v_2 . Значения этих коэффициентов вычисляются по координатам начальной точки отображаемых областей и постоянны для всей области.

Величины P_j, Q_j, P_j' и Q_j' являются функциями приращений координат, соответственно $\Delta u_1, \Delta v_2, \Delta u_1$ и Δv_1 .

Выражения (1) – (4) позволяют получить общий алгоритм вычислений для широкого спектра геодезических проекций на основе конформного отображения, служащий основой для формирования координатной среды геоинформационной системы.

Основные способы представления знаний

Отличие искусственного (формализованного) языка от естественного заключается в возможности воспроизведения логических связей между объектами предметной области в доступных для машинной обработки сжатых символьных формах. Функциональные классы естественного языка могут служить основой процесса формализации знаний.

Среди основных способов представления знаний при помощи функциональных классов естественного языка выделим RX -коды, синтагматические цепи и семантические сети [9-11].

В RX -кодах понятие определяется выражением вида

$$X_n = R_1 X_1 R_2 X_2 \dots R_k X_k, \quad (5)$$

где X_n – некоторое базовое понятие; X_i – понятие, отличное от понятия X_n ; R_i – отношения между понятиями; $i = 1, 2, 3, \dots, k$.

Пример 1. Отношения между понятиями «поверхность» и «начальная точка» можно обозначить в виде RX -кода

$$X_1 = R_1 X_2, \quad (6)$$

где X_1 – понятие «поверхность», X_2 – понятие «начальная точка», R_1 – отношение «определяется».

RX -коды легко отображаются в виде графов, что позволяет визуализировать отношения между понятиями-классами (рис. 1).

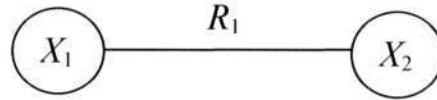


Рис. 1. Графическое представление RX -кода (6)

RX -коды позволяют использовать подстановки одних понятий вместо других. Например, выражение естественного языка «поверхность определяется начальной точкой, заданной изометрическими координатами», можно записать в виде

$$X_1 = R_1(R_2X_3), \tag{7}$$

где X_3 – «изометрические координаты», R_2 – «задана».

Пример 2. Выразим в виде RX -кода (8) и изобразим графически выражение «поверхность определяется начальной точкой, заданной изометрическими координатами, и состоит из точек поверхности, заданных изометрическими координатами» (рис.2):

$$X_1 = R_1(R_2X_3)R_3(R_2X_3) = R_1X_2R_3X_4, \tag{8}$$

где X_4 – «точки поверхности», R_3 – «состоит».

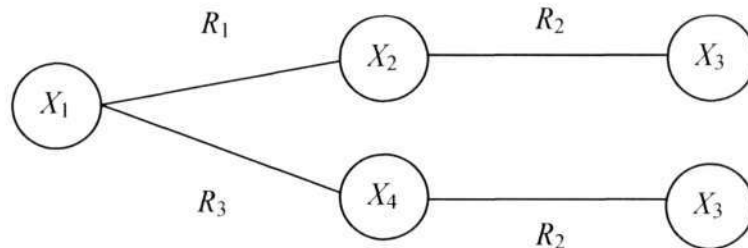


Рис. 2. Графическое представление RX -кода (8)

Представление выражений в виде RX -кодов позволяет структурировать понятия для их сравнения между собой. Главный недостаток RX -кодов – это их статичность при описании отношений между понятиями: они постоянны во времени и не зависят от контекста. Этот недостаток преодолевается представлением знаний при помощи синтагматических цепей.

Элементарной синтагмой называется тройка элементов $(x r y)$, состоящая из понятий x, y и отношения r между ними. Синтагматическая цепь образуется в процессе подстановки одной элементарной синтагмы вместо другой и связывания синтагм логическими связками и кванторами.

Выражение «точка отображаемой поверхности определяется первой и второй координатами» можно показать в виде следующей синтагматической цепи:

$$(a_1 r_1 a_2) \wedge (a_1 r_1 a_3), \tag{9}$$

где a_1 – «точка отображаемой поверхности», a_2 – «первая координата, u_2 », a_3 – «вторая координата, v_2 », \wedge – логический оператор «И».

Пример 3. Представим выражение (3) в виде однозначных синтагматических цепей (10), (11), соответствующих выражению естественного языка «приращение координат точки ото-

бражаемой поверхности есть разность значений координат точки отображаемой поверхности и значений координат начальной точки отображаемой поверхности»:

$$(a_8 r_1 a_4) \wedge (a_8 r_1 a_5), (a_4 r_1 (a_2 r_2 a_6)) \wedge (a_5 r_1 (a_3 r_2 a_7)); \quad (10)$$

$$(a_8 r_1 (a_2 r_2 a_6)) \wedge (a_8 r_1 (a_3 r_2 a_7)), \quad (11)$$

где $a_4 - \Delta u_2$; $a_5 - \Delta v_2$; $a_6 - u_{02}$; $a_7 - v_{02}$; a_8 – «приращение координат»; r_2 – «вычесть».

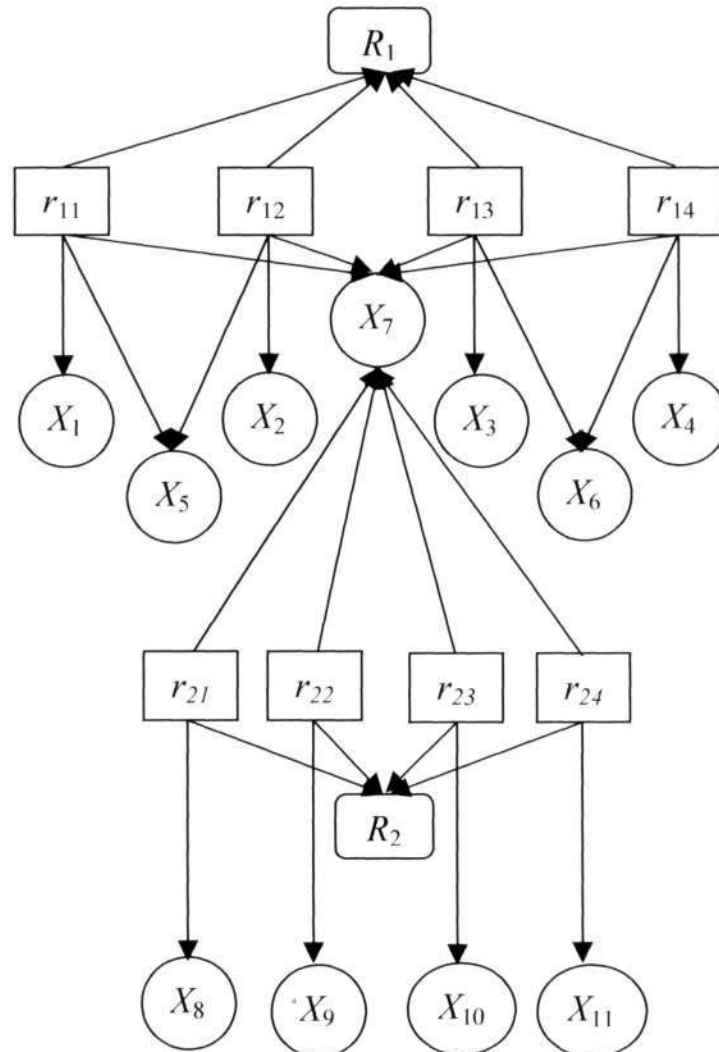


Рис. 3. Графический вариант семантической сети, отражающий закономерности отношений и связей понятий выражений (3), (4)

Выявление синтагматических отношений имеет особое значение для организации запросов пользователя (один и тот же объект может нести различную смысловую нагрузку).

RX -коды и синтагматические цепи являются частным случаем семантических сетей, задаваемых набором элементов:

- понятия (помеченная вершина, соответствующая некоторому элементу из множества X);
- отношения (вершина, соответствующая идентификатору некоторого отношения R_i);
- связи (вершина, соответствующая конкретному элементу множеств R_i).

Пример 4. Выражения (3) и (4) представим в виде семантической сети. Введем следующие обозначения понятий, отношений и связей, формализующих (3) и (4): $X_1 - C_j$; $X_2 - P_j$; $X_3 - Q_j$; $X_4 - C_j$; $X_5 - P_j$; $X_6 - Q_j$; $X_7 - j$ (степень разложения); $X_8 - C_j P_j$; $X_9 - C_j Q_j$; $X_{10} - C_j P_j$; $X_{11} - C_j Q_j$; R_1 – «выполнить произведение»; R_2 – «выполнить суммирование».

Тогда отношения R_1 и R_2 можно, соответственно, представить в виде $R_1 = \{r_{11}, r_{12}, r_{13}, r_{14}\}$, $R_2 = \{r_{21}, r_{22}, r_{23}, r_{24}\}$, где $r_{11} = \langle X_1, X_2, X_7 \rangle$; $r_{12} = \langle X_1, X_3, X_7 \rangle$; $r_{13} = \langle X_4, X_5, X_7 \rangle$; $r_{14} = \langle X_4, X_6, X_7 \rangle$; $r_{21} = \langle X_8, X_7 \rangle$; $r_{22} = \langle X_9, X_7 \rangle$; $r_{23} = \langle X_{10}, X_7 \rangle$; $r_{24} = \langle X_{11}, X_7 \rangle$.

Графический вид семантической сети, соответствующий приведенному в примере 4 набору понятий, отношений и связей, показан на рис. 3.

Заключение

Преимущества описанных способов представления знаний заключаются в их экономичности (сокращении времени автоматизированного поиска информации) и удобстве описания определенных областей знания, когда выделяются основные объекты предметной области и система понятий, в которых анализируются конкретные ситуации, описываются свойства объектов и отношения между ними [2].

Представление знаний с помощью функциональных классов естественного языка позволяет наглядно отражать взаимоотношения между понятиями конкретной предметной области в виде графов и создает основу для использования формальных языков. Более того, графические и формальные способы представления знаний, например при помощи RX -кодов, синтагматических цепей и семантических сетей, могут служить основой проектирования концептуальных схем баз данных геоинформационных систем, основанных на объектно-ориентированных или реляционных моделях данных.

Список литературы

1. Ревунков Г.И., Самохвалов Э.Н., Чистов В.В. Базы и банки данных и знаний. – М.: Высшая школа, 1992. – 367 с.
2. Алексеева И.Ю. Человеческое знание и его компьютерный образ. – М.: Институт философии РАН, 1993. – 218 с.
3. Элти Дж., Кумбс М. Экспертные системы: концепции и примеры. – М.: Финансы и статистика, 1987. – 388 с.
4. Галанский Б.Л., Поляков В.И. Информационные системы. – Томск: Изд-во Томского ун-та, 1989. – 230 с.
5. Хейес-Рот Ф., Уотерман Д., Ленат Д. Построение экспертных систем. – М.: Мир, 1987. – 422 с.
6. Смирнова Е.Д. Формализованные языки и проблемы логической семантики. – М.: Изд-во Московского ун-та, 1982. – 182 с.
7. Философский словарь / Под ред. И.Т. Фролова. 5-е изд. – М.: Политиздат, 1986. – 590 с.
8. Подшивалов В.П. Теоретические основы формирования координатной среды для геоинформационных систем. – Новополоцк, 1998. – 126 с.
9. Тельнов Ю.Ф., Скорова А.А., Андреева Н.В. Проектирование баз знаний. – М.: Эконом. ин-т, 1991. – 109 с.
10. Першиков В.И., Савинков В.М. Толковый словарь по информатике. – М.: Финансы и статистика, 1991. – 543 с.
11. Уэно Х., Исидзука М. Представление и использование знаний. – М.: Мир, 1989. – 220 с.

Поступила 06.02.04

Научно-исследовательский институт
Вооруженных Сил Республики Беларусь,
Минск, ул. Калиновского, 4

I. V. Yauhleuski

**KNOWLEDGE REPRESENTATION FOR COORDINATE ENVIRONMENT IN
GEOINFORMATION SYSTEMS BY MEANS OF FUNCTIONAL CLASSES
OF THE NATURAL LANGUAGE**

Bases of formalization of knowledge are resulted by the example of formation of the coordinate environment for geographical information systems. Functional classes of a natural language on the basis of which are formalized knowledge are described. By the example of display of surfaces process of construction of RX-codes, syntagmatic circuits and semantic networks with a view of modelling a subject domain for formation of the coordinate environment of geoinformation systems is shown.