

МОДЕЛЬ ОНТОЛОГИЧЕСКОГО ИНТЕРФЕЙСА АГРЕГАЦИИ ИНФОРМАЦИОННЫХ РЕСУРСОВ И СРЕДСТВ ГИС

Марина Попова

Аннотация: В статье рассматривается разработка и применение онтологического интерфейса как эффективного средства обеспечения процессов интеграции распределённых информационных ресурсов и систем на основе использования семантических свойств. Описано представление информации в наглядной легкодоступной форме, обеспечивающее создание и использование формализованной системы знаний в конкретных предметных областях. Приведён пример агрегации онтологического интерфейса с ГИС-средой.

Ключевые слова: онтология, онтограф, онтологический интерфейс, геоинформационные системы (ГИС).

ACM Classification Keywords: I.2 ARTIFICIAL INTELLIGENCE - I.2.4 Knowledge Representation Formalisms and Methods

Введение

Восприятие и познание окружающего мира требуют развития соответствующих методов и средств, среди которых выделяются геоинформационные системы (ГИС). ГИС могут применяться в широком спектре задач, связанных с анализом и прогнозом явлений и событий окружающего мира, с осмыслением и выделением главных факторов и причин, а также их возможных последствий, с планированием стратегических решений последствий совершаемых действий [Людвиг Фон Берталанфи, 1969]. В свою очередь, развитие геоинформационных систем связано с необходимостью совместной обработки объемов пространственной и непространственной информации, сложных процессов обработки взаимосвязанной разноплановой информации, ее интеграции и взаимодействия с другими различными по назначению системами. Дополнительные требования нахождения лучших решений, удобства, производительности, надежности и стоимости также требуют разработки и развития адекватных моделей.

Быстрый рост объемов информации, необходимость ее более качественной обработки и усвоения требуют использования методов извлечения информации и преобразования ее в такую форму, с которой будет удобно работать. Главная цель такого преобразования – возможность анализа «хаотичной» информации с помощью стандартных методов обработки данных. Более специфической целью является выявление логических закономерностей между описанными понятиями. Представленная надлежащим образом информация позволяет увидеть те дополнительные скрытые закономерности, которые не удается обнаружить другими методами.

Таким образом, актуальной является задача идентификации, поддержки и управления пространственными связями между топологическими объектами реального мира, создание новых объектов, связей, увязки новых атрибутов, визуализирующихся в виде «дружественного интерфейса».

Анализ последних исследований и публикаций

Согласно анализу современных методов и средств визуального представления информационных ресурсов в ГИС-среде [Joseph K. Berry, 2007] можно сделать вывод о том, что широко применяемым является объектный подход, при котором предметная прикладная область представляется в виде совокупности объектов, которые взаимодействуют между собой посредством передачи сообщений.

Распространенное использование ГИС создало основу для широкого применения объектно-ориентированного подхода в практике проектирования и программирования информационных систем. Указанная методология ориентирована, прежде всего, на преодоление сложностей, связанных с разработкой программных средств, созданием больших сложных систем, коллективной их разработкой, дальнейшим активным сопровождением при эксплуатации и регулярных модификациях.

Информационно-аналитическая среда ГИС-специалиста может иметь иерархическую или сетевую структуру, т.е. состоять из более специализированных сред или систем, связанных некоторыми отношениями, например, «общее-частное». Поскольку архитектура такой среды должна предоставлять гибкость, возможность для расширения функциональности и агрегации распределенных в сети информационных ресурсов, необходимо создание эргономичного интерфейса пользователя, обеспечивающего Web-доступ к ним.

Эффективным средством представления и систематизации информации являются онтологии, которые используются для формальной спецификации понятий и отношений, которые, в свою очередь, характеризуют определенную предметную область. Преимуществом онтологий как способа представления информации является их формальная структура, которая упрощает компьютерную обработку [Палагин А.В., 2005].

Будучи аналогом понятию «модель», онтология служит средством коммуникации между разработчиком и пользователем.

Использование онтологии эффективно при поиске и объединении информации из различных источников и сред, представления и интерпретации информации в процессе принятия решений. Онтологический подход обеспечивает связность информационных ресурсов и позволяет гибко работать с контекстами.

Современные интерфейсы систем представления информационных ресурсов предназначены для функционирования в гетерогенных распределенных информационных средах и поэтому основаны на методах искусственного интеллекта и парадигме Semantic Web.

На сегодняшний день одним из доминирующих решений Web-технологий, которое сводит к единой структуре как корпоративные документы и материалы, так и Интернет-ресурсы, являются порталы. Различают разные типы порталов, в зависимости от функций, которые они выполняют.

Простейший тип – информационные порталы (Information Portals), которые объединяют пользователей с информацией, обеспечивают персонализированный доступ к ресурсам и данным с помощью классификатора, с возможностью проведения сквозного полнотекстового и атрибутивного поиска.

Порталы для совместной работы (Collaboration Portals) поддерживают различные средства взаимодействия пользователей, основанные на компьютерных технологиях. Такие порталы предоставляют информацию и обеспечивают работу группы сотрудников над определенной задачей, проектом (фактически, автоматизируют бизнес-процессы в организациях).

Экспертные порталы или порталы экспертизы (Expertise Portals) объединяют пользователей друг с другом на основе их опыта, области экспертизы и интересов. Такие системы обеспечивают подключение к экспертам на основе их знаний. Не всегда такой портал содержит нужную пользователю информацию, однако, в случае ее отсутствия может подключить пользователя к соответствующим специалистам, которые могут поделиться необходимой информацией, предоставить экспертные оценки по конкретным вопросам.

Порталы знаний (Knowledge Portals) – это интегрированные порталы, которые объединяют возможности вышеперечисленных типов и обеспечивают предоставление персонализированной информации с учетом конкретной работы, которую выполняет каждый пользователь в определенное время [Шинкарук В.Д., 2008].

Все перечисленные свойства порталных решений (комбинирование в себе трёх типов порталов с использованием компьютерных технологий, соединяющих пользователей с информацией и пользователей друг с другом на основе таких критериев как опыт, область экспертизы, общие интересы и т.п.) имеет онтологический интерфейс. В основе такого интерфейса лежит онтология, которая условно делится на две части: первая содержит описание структуры ГИС-среды, вторая – ресурсы, описывающие выбранную предметную область.

Модель онтологического интерфейса

На сегодняшний день информационные ресурсы, используемые в процессе принятия решений, являются распределенными. Современные сетевые технологии и широкое распространение Internet предоставляют возможность доступа и использования этих ресурсов путем объединения территориально распределенных источников информации такого рода. Онтологический интерфейс позволяет визуализировать результат процессов интеграции и агрегации распределенных информационных ресурсов в процессе организации взаимодействия пользователей в легкодоступной наглядной форме.

Компьютерная онтология предметной области – это:

- иерархическая структура конечного множества понятий, описывающих заданную предметную область (ПДО);
- структура – онтограф, вершинами которого являются понятия, а дугами – семантические отношения между ними;
- понятия и отношения интерпретируются согласно общезначимым функциям интерпретации, взятых из электронных источников знаний заданной ПДО;
- определение понятий и отношений выполняется на основе аксиом и ограничений (правил) их области действия;
- существует средство формального описания онтографа;
- функции интерпретации и аксиомы описаны в нотации формальной теории.

Онтология определяет общеупотребительные, семантически значимые «понятийные единицы информации», которыми оперируют пользователи и разработчики информационных систем. В отличие от информации, закодированной в алгоритмах, онтология обеспечивает ее унифицированное и многократное использование различными группами пользователей, на разных компьютерных платформах при решении различных задач.

Онтология некоторой ПдО в общем случае формально представляется Т. А. Гавриловой и Ф. В. Хорошевским [Гаврилова Т.А., 2001] упорядоченной тройкой :

$$O = \langle X, R, F \rangle, \quad (1)$$

где X, R, F – конечные множества соответственно:

X - концептов (понятий, терминов) ПдО;

R – отношений между ними;

F - функций интерпретации (определений) X и/или R .

Выделяем 5 типов онтологий:

$X = \emptyset, R = \emptyset, F = \emptyset$ – неструктурированный текст;

$X \neq \emptyset, R = \emptyset, F \neq \emptyset$ – глоссарий;

$X \neq \emptyset, R \neq \emptyset, F = \emptyset$ – таксономия;

$X \neq \emptyset, R = \emptyset, F = \emptyset$ – простая онтология;

$X \neq \emptyset, R \neq \emptyset, F \neq \emptyset$ – активная онтология.

Активная онтология ($R \neq \emptyset, F \neq \emptyset$) – это такая онтология, в которой множества концептов и концептуальных отношений максимально полные, а к функциям интерпретации добавляются аксиомы, определения и ограничения. Описание всех компонент представлено некоторым формальным языком, доступным для их интерпретации компьютером.

$$O = \langle X, R, F, A(D, Rs) \rangle, \quad (2)$$

где $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$, $i = \overline{1, n}$, $n = \text{Card } X$ – конечное множество концептов (понятий-объектов) заданной ПдО;

$R = \{R_1, R_2, \dots, R_k, \dots, R_m\}$, $R \subseteq X_1 \times X_2 \times \dots \times X_n$, $k = \overline{1, m}$, $m = \text{Card } R$ – множество концептуальных отношений между ними;

$F: X \times R$ – конечное множество функций интерпретации, заданных на концептах и/или отношениях;

A – конечное множество аксиом, состоящее из множества определений D_i и множества ограничений Rs_i для понятия X_i . Определения записываются в виде тождественно истинных высказываний, которые могут быть взяты, например, из толковых словарей ПдО. В словарях могут быть указаны дополнительные взаимосвязи понятий X_i с понятиями X_j . В множестве ограничений Rs_i могут быть заданы ограничения на интерпретацию соответствующих понятий X_i ;

D – множество дополнительных определений понятий;

Rs – множество ограничений, определяющих область действия понятийных структур.

Рассмотрим множество ограничений и множество дополнительных определений:

$D = X \times R \times Rs$ – множество дополнительных определений;

$Rs = R^* \times R$ – множество ограничений, которое может быть рассмотрено как замыкание отношений R ;

R^* – множество качеств, характеризующих элементы множества R .

Поскольку любое информационное пространство представляет собой сложную систему управления взаимодействием пользователей с информационной системой, пользователей между собой, а также

является средством интеграции распределенных информационных ресурсов и процессов, то под *информационной системой* следует понимать совокупность организационных и технических средств для хранения и обработки информации с целью обеспечения информационных потребностей пользователей.

Системными компонентами являются:

- Типы данных, интерпретирующие процессы;
- Процедуры, обрабатывающие соответствующие типы данных;
- Источники, определяющие непосредственно типы данных и задающие их значение;
- Потребители или фиксирующие устройства.

Информационная система рассматривается через множество представителей - задач, которые могут быть решены с помощью информационной системы.

Задача проблемной ситуации с набором заданных целей может быть представлена в виде кортежа

$$T = \langle K, K^*, Aim \rangle, \quad (3)$$

K – модель ПдО, отображающая проблемную ситуацию;

K^* – кортеж состояний ПдО, актуализирующихся на каждом шагу достижения целей;

$$K^* = \langle K_0, K_1, \dots, K_i, \dots, K_n \rangle, \quad (4)$$

$Aim = F \times R$ – набор целей.

Таким образом, *онтологический интерфейс* имеет вид:

$$I = \langle K, K^*, F \times R, X, R, F, A, (X \times R \times R_s, R^* \times R) \rangle, \quad (5)$$

$$I = \langle K, K^*, Aim, X, R, F, A, (D, R_s) \rangle, \quad (6)$$

Онтологический интерфейс – средство удобного взаимодействия пользователя с информационной системой, предназначенной для решения множества задач проблемной ситуации путем использования активной онтологии.

$$I = \langle T, O \rangle, \quad (7)$$

Формально технологический базис формирования онтологического интерфейса определяется нагруженным двудольным графом.

$$G = (V_1 \cup V_2, E), \quad (8)$$

где $V_1 \cap V_2 = \emptyset$, вершины из V_1 размечены именами предикатов, а вершины из V_2 – именами аргументов;

E – множество дуг (рёбер). Дуги графа соединяют вершины, размеченные именами предикатов, с вершинами, размеченными именами аргументов.

Вершины из множества V_1 называются узлами-предикатами, вершины из V_2 – узлами-концептами, а сами предикаты – концептуальными сущностями.

Высказывания формируются на основе композиции вершин, инцидентных одному ребру.

Алгоритм формирования:

1. Определяется первая вершина (левая или правая) по направлению отношения, если оно не коммутативное;
2. Выбирается левая/правая вершина и инцидентное ребро;

3. Выбирается правая/левая вершина с инцидентным ребром, которое имеет левую/правую вершину;
4. Двудольный граф определяется как высказывание.

Вычисляется значение выражения: истинность – вершины включаются в множество объектов интерфейса, ложность – вершины не входят в этого множество.

Алгоритм формирования объектов онтологического интерфейса как множества истинных высказываний может быть представлен в общем виде нормального алгоритма Маркова [Марков А. А., 1996].

Визуализация информации в виде иерархического графа помогает пользователю:

- быстро находить нужный элемент в иерархии;
- понимать связь элемента с контекстом;
- обеспечивать возможность прямого доступа к информации при вершинах.

Сетевой граф может выступать не только средством организации информации. Расширяя его традиционные функции благодаря отражению в виде онтологического интерфейса, граф можно превратить в среду, в которой обеспечивается активная работа с распределенными информационными ресурсами.

Формирование онтологического интерфейса состоит из четырёх этапов:

1. Предварительный анализ информационных ресурсов по ПдО. Выделение концептов-понятий и объединение их по свойствам в соответствующие классы.
2. Формирование таблицы классов концептов-понятий на основе множества семантических соответствий между понятиями.
3. Построение онтографа.
4. Визуализация онтографа и формализованное описание онтологии [Стрижак О.Е., 2013].

На рис. 1 изображён фрагмент онтографа кристаллохимической классификации минералов «IMA CNMNC mineral classes» и соответствующий ей онтологический интерфейс. В данном примере экземпляры минералов входят в состав минерального класса «03 Halogenides», который в свою очередь входит в состав классификации «IMA CNMNC».

Онтологический интерфейс предназначен для визуального отображения экземпляров минералов и их принадлежности к классам в классификации, агрегации распределённых информационных источников по минералогии, а онтограф является таксономией и выполняет функции редактирования, дополнения, управления информационными ресурсами и выявления новых связей или объектов. Например, минерал пирит «Pyrite» входит в класс сульфидов «02 Sulfides and Sulfosalts» кристаллохимической классификации и в класс руд «Ores» классификации по общим признакам. С помощью онтографа был определён новый класс минералов, который можно описать как «рудные сульфиды» или «сульфидные руды» (рис.2).

Использование геоинформационных систем предоставляет дополнительные возможности визуализации данных с географической привязкой и использования многочисленных аналитических инструментов для повышения уровня эффективности принятия решений.

В геоинформационных системах классы объектов онтологии составляют слои тематической карты, а сами объекты, входящие в соответствующий класс, являются объектами слоя.

Атрибутивная информация об объектах онтологии, представленная в онтографе, отображается на карте в виде вложений. Каждая вершина онтографа имеет собственную «базу данных», содержащую

информацию из распределённых информационных ресурсов, необходимую для тщательного ознакомления с выбранным объектом, которая может пополняться данными и поисковыми запросами пользователей ГИС [Попова М. А., 2013].

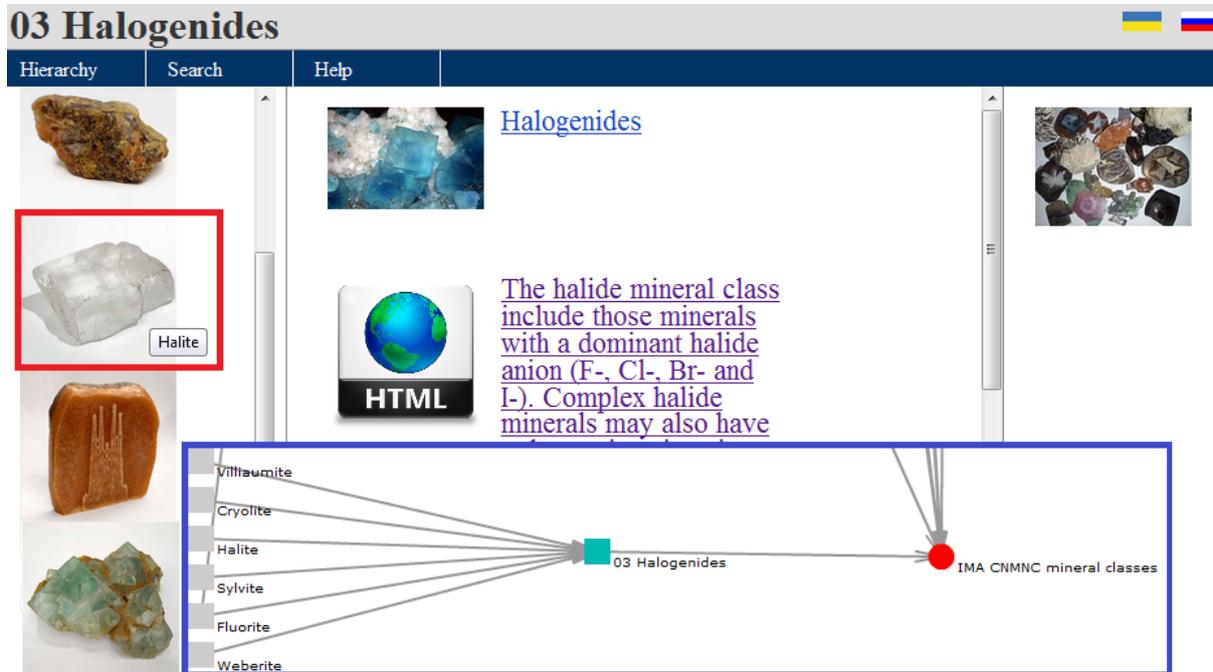


Рис. 1. Онтологический интерфейс онтографа кристаллохимической классификации минералов «IMA CNMNC mineral classes»

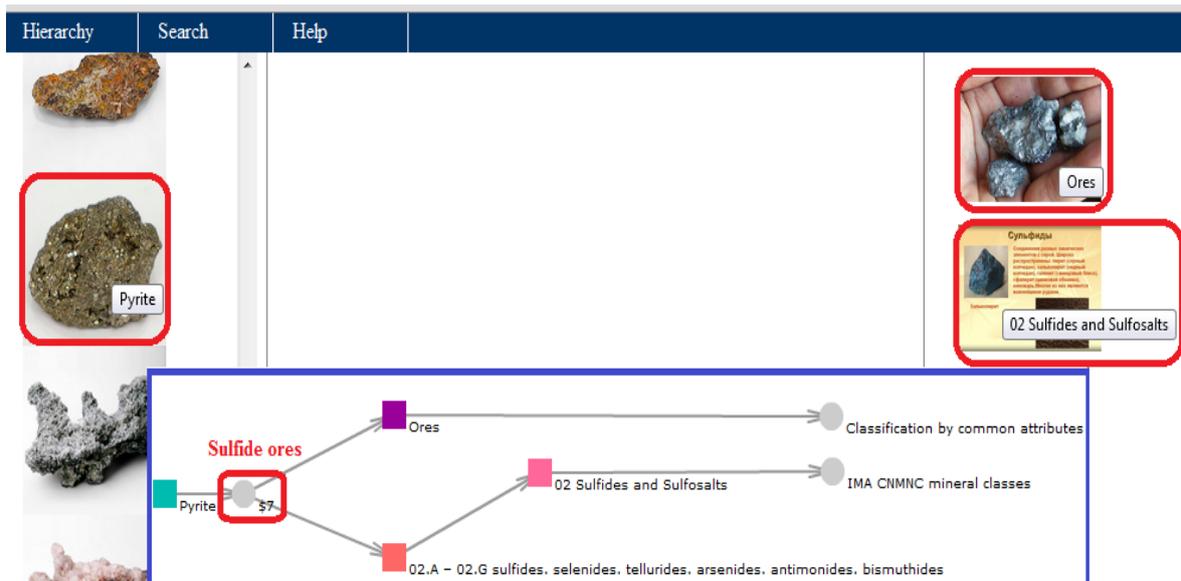


Рис. 2. Выявление нового класса объектов на примере классификаций минералов

На рис. 3 приведён пример отображения слоя «03 Halogenides» на карте, соответствующего одноименному классу онтографа «IMA CNMNC mineral classes». Все объекты, входящие в данные класс и описанные в онтографе, отображаются на карте как элементы, составляющие слой. Выбранный в онтологическом интерфейсе объект «Halite» (рис.1) наследует в ГИС-среде все свойства, заданные в онтографе.

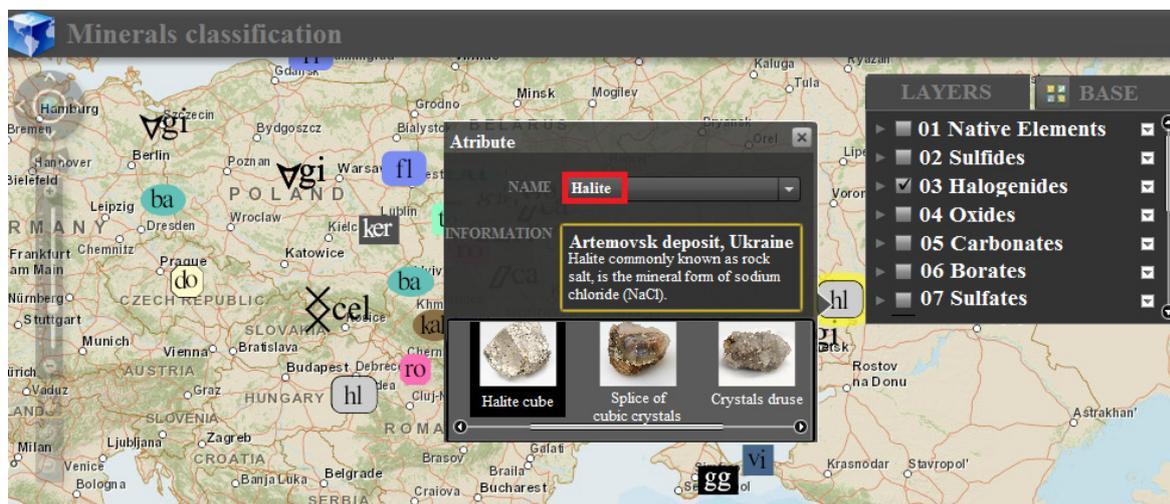


Рис. 3. Фрагмент тематической карты с отображением активного слоя класса объектов онтографа

Выводы

Использование онтологического подхода к классификации, систематизации и использованию информационных ресурсов и онтологического интерфейса для визуализации агрегации распределенных информационных моделей и систем на основе использования семантических свойств дает возможность каждому пользователю выявлять принципиально новые взаимосвязи или объекты, неизвестные ранее. Активные методы анализа проблем и поиска решений способствует смещению акцентов с пассивных методов поиска, ориентированных на передачу данных, к более широкому применению разноформатной разнородной распределённой информации в единой аналитической среде с использованием возможностей ГИС-технологий.

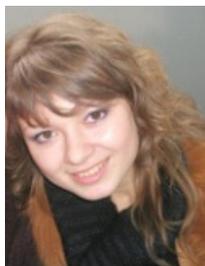
Рассмотренная модель онтологического интерфейса как множества описаний процессов активизации состояний онтологии является средством управления взаимодействия пользователей с информационной системой и друг с другом, интеграции распределённых информационных ресурсов и процессов. Необходимой является разработка методов и средств контекстного расширения модели предметной области путём интеграции онтологических интерфейсов.

Библиография

[Людвиг Фон Бертуланфи, 1969] Бертуланфи Людвиг Фон. Общая теория систем: обзор проблем и результатов/ Бертуланфи Людвиг Фон // Системные исследования. – М.: Наука, 1969. – С. 30-54.

-
- [Joseph K. Berry, 2007] Joseph K. Berry. Beyond Mapping III. Understanding Spatial Patterns and Relationships - BASIS Press, 2007, 227 p. [Electronic Resource] .- URL: <http://www.innovativegis.com/basis/MapAnalysis/>
- [Палагин А.В., 2005] Палагин А.В. Системная интеграция средств компьютерной техники / А.В. Палагин, Ю.С. Яковлев. – Винница: УНІВЕРСУМ, 2005. – 680 с.
- [Шинкарук В.Д., 2008] Шинкарук В.Д. Здійснення процедур акредитації, ліцензування та рейтингування вищих навчальних закладів України з використанням онтологічного підходу. Навчально-методичний посібник / Укладачі: В.Д. Шинкарук, М.В. Михайліченко, М.Ф. Бондаренко, Н.С. Лесна, Н.В. Рябова, М.В. Климова, Н.О. Волошина, В.Б. Репка, Г.А. Воскобойникова. – Харків: ХНУРЕ, 2008.
- [Гаврилова Т.А., 2001] Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб.: Питер, 2001. – 384 с.
- [Марков А. А., 1996] Марков А. А. Теория алгорифмов./ А. А. Марков, Н. М. Нагорный — М.: Наука, 1984. — 432 с. — (Мат. логика и основания математики). || . — 2-е изд., испр. и доп. — М.: Фазис, 1996. — 493 с. — 2000 экз. — ISBN 5-7036-0020-0
- [Стрижак О.Є., 2013] Стрижак О.Є. Засоби онтологічної інтеграції і супроводу розподілених просторових та семантичних інформаційних ресурсів. - Екологічна безпека та природокористування: Зб. наук. праць / М-во освіти і науки України, Київ. нац. ун-т буд-ва і архіт., НАН України, Ін-т телекомунікацій і глобал. інформ. простору; редкол.: О.С. Волошкіна, О.М. Трофимчук (голов. ред.) [та ін.]. — К., 2013. — Вип. 12. —1988 с.: іл. – Бібліогр. в кінці ст.
- [Попова М. А., 2013] Попова М. А. Онтологічний інтерфейс як засіб представлення інформаційних ресурсів в ГІС-середовищі / М.А. Попова, О.Є. Стрижак // Ученые записки Таврического национального университета имени В.И. Вернадского. Серия: География. – 2013. – Т. 26 (65). – № 1– С. 127-135.

Сведения об авторе



Марина Попова – младший научный сотрудник Института телекоммуникаций и глобального информационного пространства НАН Украины, 03186 Украина, Киев, Чоколовский бульвар 13;

e-mail: pma1701@gmail.com