

ПОСТРОЕНИЕ ТАКСОНОМИИ ДОКУМЕНТОВ ДЛЯ ФОРМИРОВАНИЯ ИЕРАРХИЧЕСКИХ СЛОЕВ В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ

Виталий Величко, Виталий Приходнюк, Александр Стрижак,
Крассимир Марков, Крассимира Иванова, Стефан Карастанев

Аннотация: Рассмотрены проблемы, связанные с использованием функциональности онтологических систем для обработки документов, представленных в глобальной сетевой среде. Описаны механизмы аппликативной типизации выражений, составляющих тексты документов на основе использования безтиповых инструментов теории лямбда-исчисления. Изложено использование отношения частичной упорядоченности для формирования таксономий документов на основе формирования безтиповых выражений из термов, задаваемых концептами, определяющих понятийную структуру документов. Описывается процедура лексического анализа текстов и построения таксономической структуры на основе применения процедуры анализа. Предложена технология формирования иерархий тематических слоев в геоинформационных системах (ГИС) на основе таксономий сетевых документов. Рассматривается возможность применения изложенного подхода к обработке пространственных данных в соответствии с директивой EC INSPIRE.

Ключевые слова: сетевые информационные ресурсы, онтология, таксономия, упорядоченность, безтиповое выражение, аппликация, лексема, лексический анализ, функция.

АСМ классификация: I.2 ARTIFICIAL INTELLIGENCE - I.2.4 Knowledge Representation Formalisms and Methods, H. Information Systems – H.2 DATABASE MANAGEMENT – H.2.4 Systems

Введение

Неструктурированная и слабоструктурированная информация, хранящаяся в глобальной сетевой среде, постоянно накапливается и архивируется [Report, 2012]. Сетевые информационные ресурсы (СИР) отображают достаточно большую совокупность тематических знаний, технологических решений, опосредованных данных, отражающих развитие современной науки и техники, состояние информационных процессов в сети и тем самым определяют контекстное содержание информационного пространства (Information networks space). Тематическое многообразие и многоаспектность сетевой информации ставит достаточно существенную проблему ее оперативного и достаточно корректного практического использования и интерпретации. При этом практически постоянно возникает необходимость в реализации и обеспечении взаимодействия сетевых информационных систем, которые характеризуются

междисциплинарностью и реализованы на основе различных технологий и стандартов [Стрижак, 2015].

Эффективность процесса использования информационных ресурсов глобальной сети при организации взаимодействия между сетевыми информационными системами, существенно зависит от операциональных характеристик инструментов (сетевых программных средств), которые обеспечивают получение, обработку и хранение соответствующего контента. И здесь приходится учитывать достаточно существенное технологическое ограничение на организацию межсистемного взаимодействия, связанную с интероперабельностью, как сетевых информационных ресурсов, определяющихся разными форматами представления, так собственно и информационных систем сети, которые, как уже было отмечено, реализуются на основе использования разных технологических стандартов [Глушков, 1982], [Белов, 2005]. Контент глобального информационного пространства определяется контекстами объектов, которые определяют тематический характер сетевой среды. На этом основании можно сформулировать гипотезу, что эффективность интегрированного использования информационных систем в глобальной сети существенно зависит от технологических возможностей операционального связывания контекстов объектов, которые участвуют в сетевом взаимодействии и определяют информационный контент собственно сетевой среды.

Для обеспечения тематической связности объектов и процессов в глобальной сети, необходимо, на наш взгляд, обеспечить формирование концептуального отображения контекстных тематик используемых информационных ресурсов и систем. Одним из технологических подходов для решения указанной проблемы контекстной связности может быть использование трансдисциплинарных онтологий, обеспечивающих моделирование пассивных и активных сетевых информационных процессов накопления, обработки, отображения и взаимодействия [Стрижак, 2015], [Палагин, 2012], [Guarino, 1994], [Гладун, 1994].

Целью статьи является определение механизмов формирования таксономий сетевых документов как технологической основы интеграции информационных ресурсов и систем глобальной сети.

Аппликативные механизмы таксономизации

Использование трансдисциплинарных онтологических систем для отображения семантики СИР [Глушков, 1982], [Палагин, 2012] в виде таксономических (иерархических) структур [Гладун, 1994], над которыми задается определенная расширяемая аксиоматика и между которыми определяются множества отношений, позволяет частично решить проблему корректной интерпретации использования онтологии при решении сложных прикладных задач. Одной из системных компонент онтологической системы является таксономия [Глушков, 1982], что отражает определенную иерархию взаимодействия концептов. При этом собственно иерархия задается с помощью бинарных отношений, определяющих характер взаимодействия между

концептами онтологии. На основании указанных отношений, определяющих взаимодействие между концептами, реализуется процедура разбиения множества концептов онтологии на классы [Буч, 1992]. Для построения онтологических классификаторов конкретных предметных областей (ПрО) на основе сложившихся классов используются таксономические категории \check{T} . Список указанных категорий определяется в процессе решения прикладной задачи на основе свойств, которые объединяют концепты по тематическим признакам [Шаталкин, 2012].

Упомянутая категория сетевых информационных ресурсов обычно представляется в виде текстов. Множества терминов-концептов X из вышеупомянутых текстов связаны между собой разными множествами семантических отношений R_{sem} . Над отношениями из каждого такого множества R_{sem} может быть задано отношение частичного порядка \tilde{p} [Малишевский, 1998]. Это отношение позволяет формулировать множество правил Rul , на основании применения которых мы можем формулировать истинные утверждения из концептов ПрО. При этом будем считать, что концепт является свободным, если он не связан с другими концептами из ПрО никакими типами отношениями. Если же между концептами установлено какое-либо отношение из множества R_{sem} , то тогда такие концепты будем определять как связанные. Тогда утверждения образуются связанными концептами, а корректность определения над ними множественного отношения частичного порядка \tilde{p} позволяет считать их выполнимыми или истинными [Клини, 1957].

Построение утверждений из концептов реализуется на основании конечных наборов правил из множества Rul , которые определяют порядок применения множественного отношения частичного порядка \tilde{p} , как над концептами из множества X , так и над семантическими отношениями из множества R_{sem} . Также отношение \tilde{p} позволяет нам формировать из концептов терминополья [Коршунова, 2009] в виде иерархических структур, где между концептами задается множественное бинарное отношение частичного порядка, которое, по сути, полностью эквивалентно отношению \tilde{p} .

Такие правила построения утверждений носят аппликативную форму и могут быть представлены в виде безтипового выражения вида [Барендрегт, 1985], [Стрижак, 2014]:

$$f_a = (\lambda x.t(x)) a = t(a) \quad (1)$$

где: λ - теория – лямбда исчисление; запись λx говорит, что это λ - терм;

X – переменная, принимающая значения из множества концептов X ;

t – выражение, которое может содержать переменную x ;

a – аргумент функции, определяющий возможные значения переменной x ;

f_a – функция, применимая к аргументу a .

Как можем видеть из (1), переменные позволяют определить свойство экстенциональности множества концептов, которое также является финитным [Загорюлько, 2008], т. к. фактически

экстенциональность концепта как определенного терма - теории определяется его принадлежностью к определенному классу. Тогда мы всегда можем ограничить истинность утверждения, введя аксиому исключаящую истинность утверждения для концептов, не входящих в один класс. Тогда правила вида (1), включающие в себя в качестве переменных и аргументов отношения из множества R_{sem} и множественное отношение \tilde{r} , обладают свойством интенциональности. Т.е. утверждения, обладающие свойством истинности, обладают еще набором других свойств, определяющих применимость данного утверждения.

Достоинством данного аппликативного выражения является то, что в качестве переменной и аргумента кроме концептов, могут быть отношения, т.е. мы можем строить правила, которые связывают между собой и концепты, и отношения.

Безтиповость позволяет нам поднять уровень абстракции только до рассмотрения истинности утверждений, сформулированных на основе применения правил онтологии ПрО.

Таким образом, каждая онтология представляет собой сложный объект со своей структурой и функциональностью. Любая каноническая форма онтологии [Гаврилова, 2000] отображает множество концептов ПрО, множество отношений между ними и множество интерпретирующих функций, применение которых обеспечивает взаимодействие различных состояний онтологии. Множества концептов определяют терминополье [Коршунова, 2009] ПрО, которое имеет семантическое представление в виде определенных утверждений. Все утверждения, построенные из концептов терминополья, могут обладать свойством истинности, при условии, что они связаны отношениями, которые корректно связывают между собой контексты каждого понятия.

Над всеми концептами, из которых могут быть построены утверждения онтологии ПрО, всегда определяется множественное отношение частичного порядка – \tilde{r} [Малишевский, 1998]. Причем это отношение также участвует в построении истинных утверждений из концептов. Если сами концепты могут конструироваться из конечного множества термов, то все утверждения, образованные при помощи λx -термов, имеют конечную длину, максимальный путь которого не будет превышать длины последовательности 2^x . Утверждения, которые обладают свойством истинности в рамках λ -исчисления, будем называть разрешимыми [Барендрегт, 1985]. Выражение (1), при условии разрешимости, также обладает свойствами монотонности, обратимости, наследственности, аддитивности и конвертируемости.

Монотонность λx -термов может быть представлена в виде $\lambda x_1 \dots x_n \dots x_j L_1 \dots L_m$, ($L - \lambda$ -терм) где каждый дополнительный терм приписывается на основе правила применения множественного отношения бинарного частичной упорядоченности. Это позволяет строить цепочки утверждений на основе операции приписывания справа нового терма. Из сказанного можно сделать еще один вывод – если концепты онтологии обладают хотя-бы одним общим свойством, или хотя-бы бинарным отношением частичного порядка, то из них можно построить разрешимое утверждение в терминах λ -исчисления.

Обратимость определяет возможность строить разрешимые утверждения из отрицания существующих термов на основе изменения множественного бинарного отношения частичной упорядоченности. Т. е. справедливо отображение вида:

$$\lambda x_1 \dots x_n \dots x_i L_1 \dots L_m \xrightarrow{\tilde{p}} \lambda x_i \dots x_n \dots x_1 L_m \dots L_1 \quad (2)$$

Более компактно в терминах λ -исчисления это выражение можно представить следующим образом. Пусть M – λ -терм и K – λ -терм, тогда их композиция « \circ » представима в виде:

$$M \circ K = K \circ M = \lambda x. x \quad (3)$$

Наследственность может быть определена и впоследствии выражена как слабая и/или как сильная. Условие слабой наследственности заключается в сохранении свойства экстенциональности λ -термом, составленного из всех концептов, обладающих этим свойством и которые образуют разрешимое утверждение. Условие сильной наследственности требует сохранения свойств экстенциональности и интенциональности, как для λ -терма, полученного в результате композиции определяющих разрешимых утверждений, так и для обратимого λ -терма.

Свойство аддитивности выводимо из самого определения λ -терма (1). Каждый концепт и отношение, составляющие разрешимое утверждение, задаются в λ -терме соответствующей переменной или аргументом, и так как каждый из них определяет конкретную часть выражения, то в результате формулируется разрешимое утверждение.

Конвертируемость определяется возможностью формирования утверждений из концептов иерархически связанных классов. Тогда в структуре сложного λ -терма вида (2), сформированного из соответствующих термов, возможна подстановка одного выражения вместо другого. Основным условием такой подстановки является существование между концептами указанных выражений, множественного отношения частичного порядка \tilde{p} .

Эффективная сетевая обработка документов, содержащих большое количество информации, реализуется на основе таксономизации их содержания, с последующим представлением в виде онтологий [Палагин, 2012],[Гладун, 1994]. Таксономизация достигается за счет применения правил из множества R_{ul} к концептам множества X , определяющих содержание конкретного документа. Для этого из множества концептов X выделяются терминополья [Коршунова, 2009], концепты которых связываются между собой отношением частичного порядка \tilde{p} . Далее реализуется процедура построения цепочек утверждений из концептов терминопольей на основе операции приписывания справа нового термина из множества X в виде безтипового терма. Это достигается за счет наличия у них общих семантических отношений из множества R_{sem} . Т.е. все концепты из множества X могут быть представлены в виде выражений вида (2)–(3) – $M \circ K = K \circ M = \lambda x. x$. При этом все выражения вида (2)–(3) содержат в себе термы, связанные между собой множественным бинарным отношением частичного порядка \tilde{p} [Шаталкин, 2012].

Обобщенная процедура формирования таксономии документов

Практически любой текстовый документ может быть представлен в виде высказываний и утверждений, истинность которых определяется применимостью концептов их составляющих, на основе определяющих каждый концепт контекстов. Т.е. фактически все концепты, составляющие СИР, могут быть выделены в определенное множество терминополь [Коршунова, 2009] и, исходя из синтаксиса самих высказываний и утверждений, представлены в виде $\langle X, R_{synt} \rangle$, где X – множество концептов терминополь, а R_{synt} – множество синтаксических отношений между концептами, определяющих набор синтаксических правил из множества R_{ul} построения истинных утверждений из выбранных концептов терминополь глобальной среды.

Сетевые инструменты выделения соответствующих концептов на основе механизмов аппликативного представления текстов документов в виде утверждений реализуются на основе применения IT-технологии ТОДОС [Величко, 2009], [Величко, 2014]. Первоначальную основу такого представления составляют механизмы индуктивного формирования пирамидальной сети из концептов, определяющих все контексты документа [Гладун, 1994], [Величко, 2009], [Величко, 2014]. Словоформы и синтаксические конструкции текста из аппликативной формы вида $L \circ R_{synt} \equiv \lambda x_1 \dots x_n$ преобразуется в аппликативную форму $X \circ R_{sem} \equiv \lambda x_1 \dots x_n$. При этом отображении всевозможные пары (x_i, x_j) , определяющие отношение r_k из множества R_{synt} , заменяются парами (x'_i, x'_j) , которые определяют семантические отношения, такие как: «входит в», «состоит из», «класс-подкласс» или «класс-объект» из множества отношений R_{sem} . Полученное аппликативное выражение $X \circ R_{sem} \equiv \lambda x_1 \dots x_n$ позволяет сформировать термы, которые состоят из концептов множества $\tilde{X} \subset X$ и связаны между собой множественным отношением частичного порядка \tilde{p} . В результате получается множество концептов, связанное отношениями следования $\langle \tilde{X}, R_{ord} \rangle$, которое и определяет первичное таксономическое отображение концептов из отобранных терминополь.

Данный процесс выполняется с помощью следующих функций:

$$c_{a,b} = (\lambda x, y. t(x, y)) a, b \equiv a \text{ имеет свойство } b \quad (4)$$

$$r_{a,b,c} = (\lambda x, y, z. t(x, y, z)) a, b, c \equiv \text{ между } a \text{ и } b \text{ есть отношение } c \quad (5)$$

На этапе выделения множеств терминополь сетевыми инструментами IT-ТОДОС обеспечивается выделение из текста документа терминов-концептов, а также определение иерархических семантических отношений между ними. Таким образом, аппликативная форма $X \circ R_{sem} \equiv \lambda x_1 \dots x_n$ представляет таксономию документа, на основании обработки которого было выделено терминополь \tilde{X} , из концептов которого формируются термы аппликативных

выражений вида $\lambda x_1, x_2, \dots, x_n. c_{x_1 p_1} \& c_{x_2 p_2} \& \dots c_{x_n p_n}$. Приписывание к нему справа аппликативного выражения вида $\lambda r_{x_1 x_2 k_{12}} \& r_{x_1 x_3 k_{13}} \& \dots r_{x_{n-1} x_n k_{n-1n}}$ преобразует оба выражения в логическую функцию, являющейся тавтологией при всех значениях их термов.

Идентификация указанных отношений R_{sem} между концептами из множества $\tilde{X} \subset X$, выполняется на основании словарей ключевых слов. Для этого используется дополнительная функция:

$$\bar{c}_{a,b} = (\lambda x, y. t(x, y)) a, b \equiv a = b \tag{6}$$

Функция для выделения конкретных концептов имеет следующий вид:

$$\lambda x_1, y, x_2. c_{x_1 p_1} \& \bar{c}_{y p} \& c_{x_2 p_2} \tag{7}$$

Определение иерархических связей R_n между концептами множества \tilde{X} реализуется на основе применения аппликативной функции вида (6), у которой входные термы представимы в форме вида: $x_1, x_2, \dots, x_n, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$. На этапе формирования соответствующего класса из указанной последовательности термов уточняется тип отношения между концептами-терминами, определяющими логическую функцию вида (4)–(7), являющейся тавтологией при всех значениях составляющих ее термов.

Обобщенная процедура поддержки процессов формирования таксономий контекстов на основе обработки сетевых документов в среде ИТ-ТОДОС приводится на рис.1.

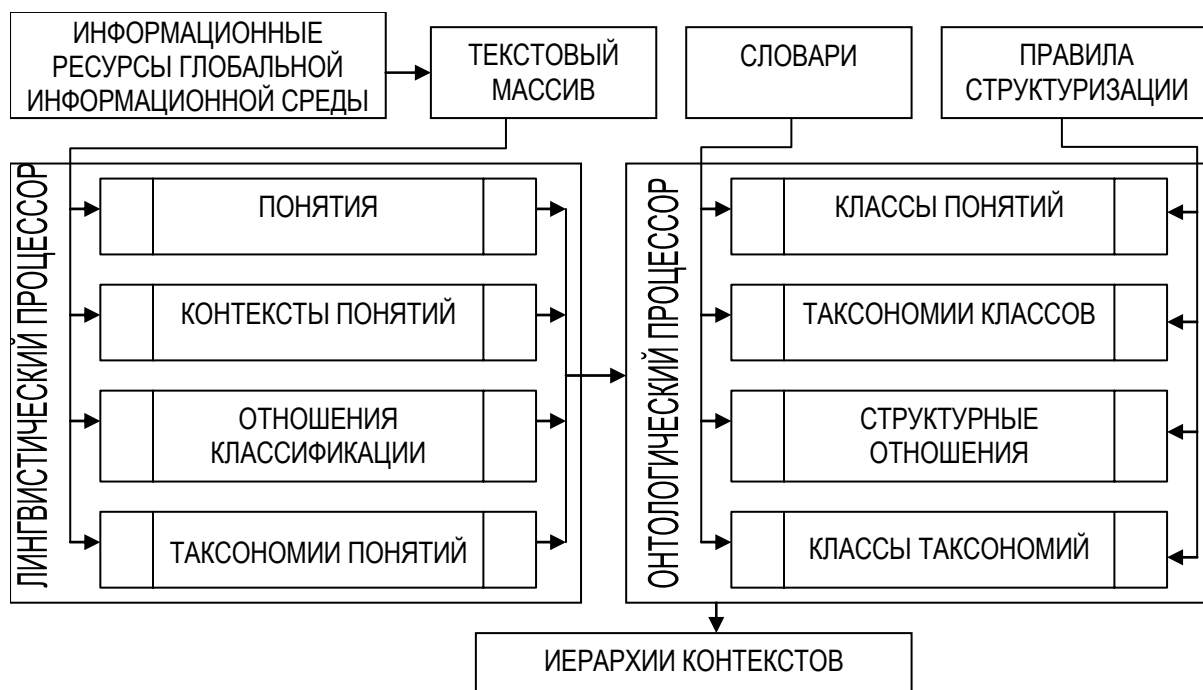


Рис. 1. Процесс формирования таксономий документов в среде ИТ-ТОДОС

Обеспечение взаимодействия в сетевой среде на основе формирования таксономий

Формирование таксономических структур, отображающих семантические свойства СИР и документов, составляющих их содержание, позволяет отобразить информационные процессы, активированные в глобальной среде, в виде безтиповых аппликативных выражений, каждое из которых представимо в виде онтологического графа [Гладун, 1994], [Палагин, 2012], [Величко, 2009]. Возникающая здесь проблема интероперабельности взаимодействия между компонентами СИР может быть сведена к определению множества отношений между терминами, составляющих множество логических функций вида (4)–(7). Все функции указанного вида являются тавтологиями. Множество термов, определяющих синтаксическую конструкцию данных функций изоморфно множеству концептов, значения которых обрабатываются сетевыми информационными системами. Таким образом, отношения между терминами, представленными в виде безтипового выражения вида (1)–(8), определяют выполнимость операций в сетевой среде над связанными компонентами СИР.

Каждый сетевой документ и каждая информационная система глобальной сети могут быть представлены в виде таксономии, поэтому процесс взаимодействия между ними сводится к установлению отношений между их состояниями при смене активных концептов или отношений между концептами. Тогда проблема интероперабельности может быть разрешима на основе применения рекурсивного редуктора вида [Барендрегт, 1985], [Стрижак, 2015]:

$$(r_{order} \rightarrow \tilde{p}) \Rightarrow ((X, R, Rul) \rightarrow L_i | i = \overline{1, n}) \quad (8)$$

Рекурсивный редуктор может быть сконструирован на основе следующих выражений:

$$r_{order} = \{(\lambda x.L_1)L_2, L_3 [x := L_2]\}, \quad (9)$$

$$Pr(x_1, \dots, x_n) = \begin{cases} 1, \neg Pr(Y\tilde{p}x_i) \wedge Pr(x_1, \dots, x_n) \\ 0, Pr(Y\tilde{p}x_i) \end{cases} \quad (10)$$

где $x_i \in X; 1 \leq i \leq n$, X, Y – множества концептов,

$Pr(x_1, \dots, x_n)$ – рекурсивный предикат [Клини, 1957].

Каждый терм-концепт таксономии может быть представлен в виде предикативного выражения

$$x = P(r_1, r_2, \dots, r_j) \quad (11)$$

При этом следует учесть, что рекурсивный редуктор вида (8)–(10) сохраняет отношение эквивалентности между таксономиями всех компонентов СИР, как документов, так и информационных систем, за счет чего и обеспечивается разрешимость проблемы интероперабельности. Таксономия может быть определена для любого сложного термина-концепта, то есть по умолчанию она может быть образована бинарным отношением линейного порядка «группа объектов - объект». Отношение «группа объектов - объект» может быть

расширено до множественного отношения «часть - целое» и далее до «быть элементом класса» и/или «быть элементом категории». Различать понятия «класс» и «категория» будем на основе полноты отображения предметной области. Определим понятие «категория» как достаточно полно отражающее семантические свойства предметной области, одновременно понятие «класс» определяется просто как отобранное множество концептов с общими семантическими свойствами.

Фактически применение предиката вида (10) совместно с выражением (11) позволяет определить, какие концепты таксономии сетевого компонента обладают данной свойством. Сформировав множество классов концептов таксономии с помощью выражений вида (8) - (10), мы получаем таксономию, над концептами которой задано множественное бинарное отношение «группа объектов - объект». Такая таксономия может иметь сложную структуру нисходящих иерархий. Каждая составляющая представляет собой класс термов-концептов, имеющих как минимум одно общее семантическое свойство.

$$\text{Pr}(x_1, \dots, x_n) = 0 \Rightarrow \exists T \subseteq \tilde{T} : \forall x \in X \exists Y \subseteq X : T = Y \tilde{r} x \quad (12)$$

Множественное отношение частичного порядка \tilde{r} , на основании применимости выражений (1)–(10) ко всем компонентам СИР, определяет взаимодействие между концептами каждой таксономии, выделенной из каждого документа или информационной системы, составляющих СИР. Соответственно применение рекурсивного редуктора (8)–(10) обеспечивает формирование новых видов таксономий, из операционально используемых на всех этапах взаимодействия компонентов, составляющих СИР.

Тогда выражения вида (1)–(12), включающие в себя в качестве переменных и аргументов отношения из множества R_{sem} и множественное отношение \tilde{r} , обладают свойством интенциональности, т. е. утверждения, формируемые из связанных общими отношениями термов-концептов, всегда обладают свойством истинности и еще набором других свойств, определяющих применимость данного утверждения.

Достоинством описанных аппликативных и предикатных выражений является то, что в качестве переменной и аргумента кроме концептов, могут быть отношения. Т. е. мы можем строить правила, которые связывают между собой и концепты, и отношения.

Безтиповость позволяет поднять уровень абстракции только до рассмотрения истинности утверждений, сформулированных на основе применения правил онтологии ПрО к таксономиям всех компонентов составляющих операциональное поле СИР.

Формирование иерархий тематических слоев в ГИС-системах на основе таксономий сетевых документов.

Одним из примеров обеспечения взаимодействия между документами и информационными системами в глобальной сети может служить формирование тематических слоев и их

иерархической взаимосвязи в среде геоинформационных систем. На основе выделенных из сетевых документов таксономий, в среде ГИС могут быть сформированы соответствующие слои тематических карт. Каждый слой строится на основе класса, представляющего объекты таксономии отобранных документов и/или сетевой информационной системы. Сами объекты, которые входят в соответствующий класс, становятся объектами соответствующего тематического слоя.

Как уже было описано, таксономия формируется на основе установления отношений между концептами и классами вида «группа объектов - объект». При этом контексты концептов, составляющих таксономию, определяют легенду карты, определяющей тематические применения ГИС. Описание объекта на карте ограничено полями атрибутивной информации, а сервис информационных приложений позволяет определить и обеспечить доступ только до такого информационного массива, который физически присутствует у пользователя. Благодаря объединению различных типов баз данных в таксономии, атрибуты объектов могут быть представлены не только в табличном виде, но и в текстовом, а также в виде гиперссылок на соответствующие документы и информационные системы глобальной сети.

Рассмотрим процесс формирования тематических слоев минералов на основании их принятой классификации, которая описывается в определенном документе. Фрагмент такого документа, в котором описываются классы минералов и составляющие их объекты, представляется в виде таблицы 1. В соответствии с таксономическим представлением, как классы объектов, так и сами объекты, а также их свойства являются термами, которые могут быть представлены в виде выражений (11)–(12). При этом выражения (8)–(12) определяют процедуру формирования собственно таксономии.

Весь документ, включая его фрагмент в виде таблицы 1, представим в виде выражения (3) – $M \circ K = K \circ M = \lambda x.x$, где каждый сложный терм вида K, M может быть представлен в виде λ -последовательности вида (2). Выражение (2) также задает процедуру формирования всех классов минералов, как концептов редуцируемой таксономии. Каждая строка таблицы документа является классом определенных концептов минералогии и может быть представлена в виде сложного λ -терма L_m . Концепт-объект класса является простым λ -термом вида λx . Принадлежность объектов к классам определяется научной классификацией минералов и позволяет выделить соответствующие свойства каждого класса.

Таким образом, получаем из документа таблицу 1, которая сформирована на основе правил (8)–(12) множества Rul .

Применение к термам выражения (3) правил (8)–(12) позволяет из термов таблицы 1 построить таксономию, представленную на рис. 2. Отметим, что одним значением некоторых термов - выражений вида (2)–(3) - могут быть географические координаты, определяющие положение каждого концепта - минерала на географической карте.

Таблица 1. Классы и концепты-объекты минералов

	Class	Relation	Object	Object	Object	Object	Object	Object	Object	Object	Object
1	Halogenides	Belongs to class	Villiaumite	Cryolite	Halite	Sylvite	Fluorite	Weberite			
2	Organic Compounds	Belongs to class	Kerrite	Shungite	Succinite	Amber	Jet (lignite)	Coral			
3	Gemstone	Belongs to class	Turquoise	Malachite	Ruby	Opal	Obsidian	Citrine	Chrysoptase	Rauhtopaz	Amethyst
4	Ores	Belongs to class	Scheelite	Zircon	Rare earth element	Copper ore	Iron ore	Complex ore	Sphalerite	Covellite	Argentite
5	Single crystal	Belongs to class	Heliodor	Rhodonite	Topaz	Labradorite	Almandine	Phlogopite	Muscovite	Natrolite	Quartz
6	Native Elements	Belongs to class	Native cobalt	Native nickel	Native copper	Native silver	Native tin	Native gold	Native platinum	Diamond	Native sulphur
7	Oxides	Belongs to class	Smokey quartz	Hawk's eye	Tiger's-cat's eye	Cat's eye	Petrified wood	Flint	Chloropal	Spinel	Corundum
8	IMA CNMNC mineral classes	Classes	Native Elements	Halogenides	Oxides	Carbonates and Nitrates	Borates	Sulfates. Selenates. Tellurates	Phosphates. Arsenates. Vanadates	Silicates and Germanates	Organic Compounds
9	Classification by common attributes	Classes	Gemstone	Single crystal	Ores						

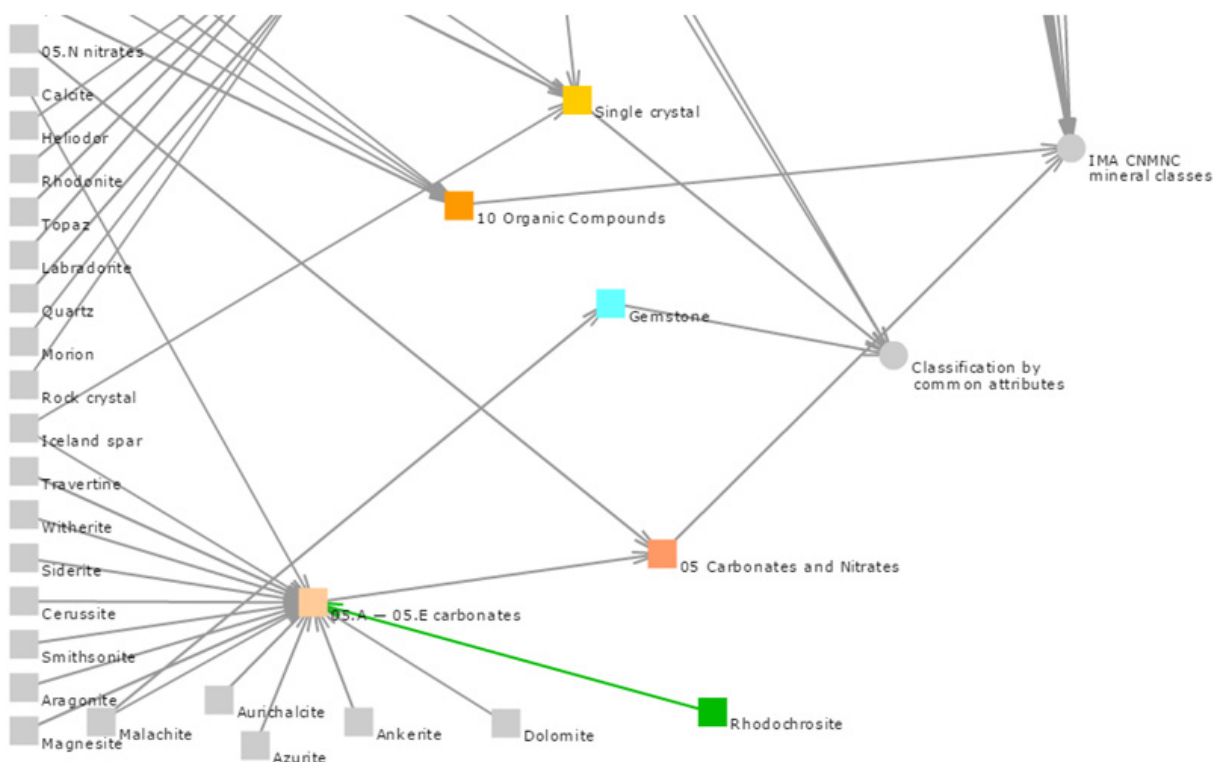


Рис.2. Фрагмент таксономии минералов

Согласно выражений (4)–(7), таксономия, представленная в виде графа (рис.2), может быть наложена на географическую карту, т.к. объекты таксономии графа имеют общее свойство с координатной сеткой карты. Т.е. описание концептов такого документа как географическая карта, а также сервисы соответствующей ГИС, могут быть включены в выражения (2)–(3). Тогда к ним применимы все правила вида (8)–(12). Это позволяет представить все концепты документа, описывающего минералы (Таблица 1) в среде ГИС (Рис.3)–(Рис 4).

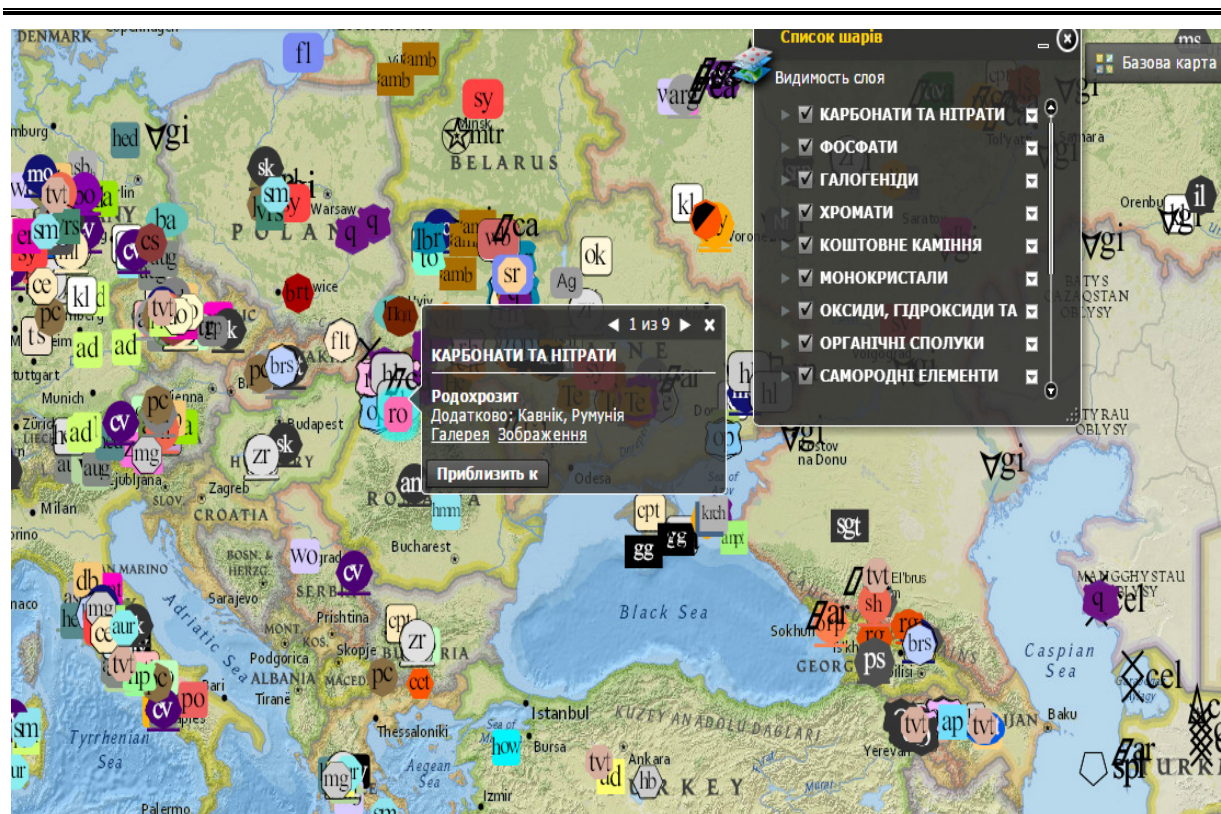


Рис.3. ГИС минералы

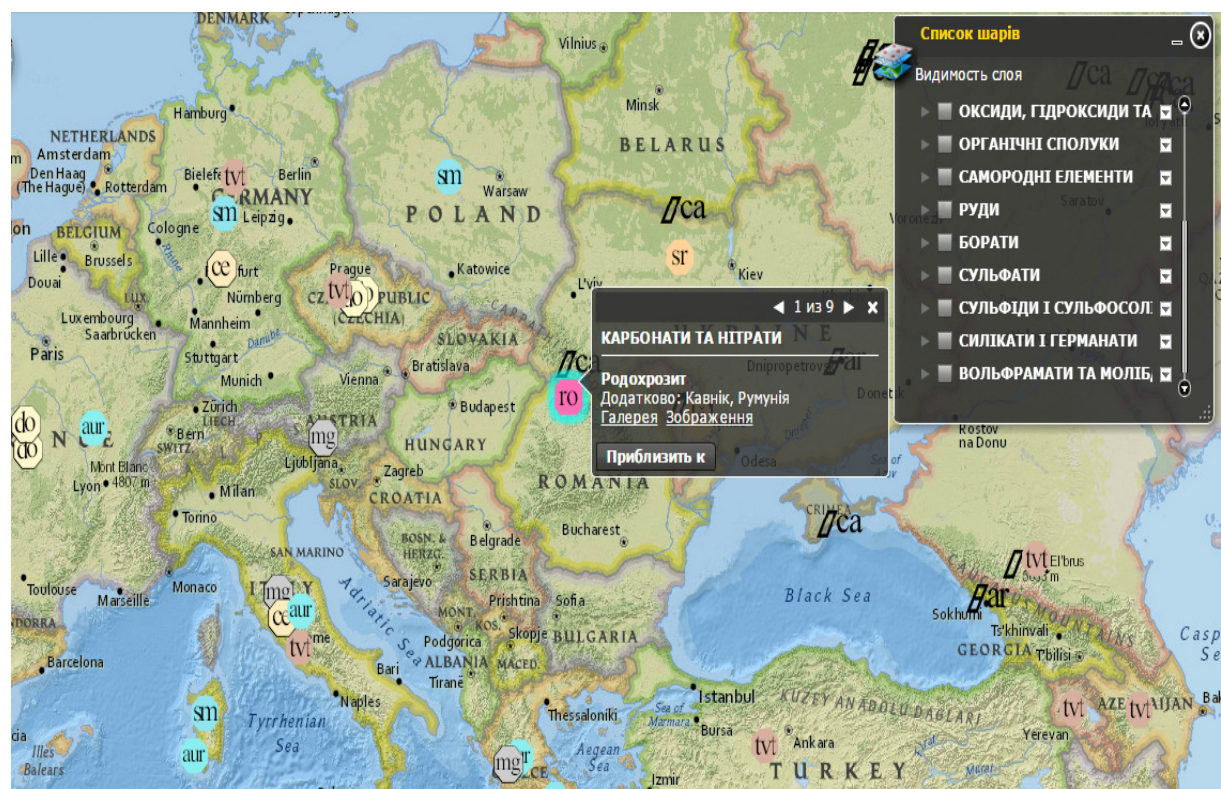


Рис.4 Фильтрация слоев карты для класса КАРБОНАТЫ и НИТРАТЫ

На основании декларирования принадлежности концепта к классу таксономии устанавливается его принадлежность к тематическому слою карты. Тогда множество всех допустимых тавтологий вида (4)–(6) состоит из утверждений о принадлежности концептов-понятий к определенной таксономии. Тавтологии формируются на основе связывания концептов классов, имеющих имена приведенных выше таксономий (Таблица 1 и Рис. 3). Как мы видим на Рис.3–4, легенда карты включает в себя тематические слои, аналогичные по имени, как классам понятий таксономии документа, так и объекты слоев, аналогичные концептам-понятиям таксономии документа.

Пространственные данные

„Пространственные данные“ представляют собой данные с прямой или косвенной ссылкой на определенную точку на земной поверхности или в географической области [INSPIRE Directive, 2007]. Рассмотренные выше данные о минералах привязаны к географическим координатам месторождений. Иными словами, данные о минералах, привязанные к точке на земной поверхности, являются „пространственными данными“, из чего следует, что изложенный выше подход возможно применить ко всем видам пространственных данных.

Директива 2007/2/ЕС Европейского Парламента и Совета от 14 марта 2007 года о создании инфраструктуры для пространственной информации в Европейском сообществе (INSPIRE) была опубликована в Официальном журнале Европейского Союза 25 апреля 2007 года и вступила в силу 15 мая 2007 года [INSPIRE Directive, 2007]. Основная цель директивы состоит в том, чтобы создать новый единый подход для обработки пространственных данных во всех странах-членах ЕС.

Упрощенный вид обычной обработки пространственных данных показан на рисунке 5. В большинстве случаев, каждая страна-член ЕС использует входные данные в соответствии с различными, часто не документированными или плохо документированными спецификациями данных. Для обработки входных данных также используются различные методы с целью получения более или менее похожей информации, имеющей отношение к деятельности в рамках ЕС [INSPIRE-DSM, 2007].

Методология, описанная в директиве INSPIRE, направлена на лучшее понимание общих требований пользователей к данным. Исходя из этих требований, методология сосредоточена на разработке согласованных спецификаций входных данных, построенных таким образом, таких чтобы все входные данные из разных стран-членов имели одинаковые характеристики и, в принципе, могли бы использовать одни и те же этапы обработки для получения необходимой информации. Обновленная схема обработки данных, разработанная с использованием предлагаемой методики, показана на рисунке 6 [INSPIRE-DSM, 2007].

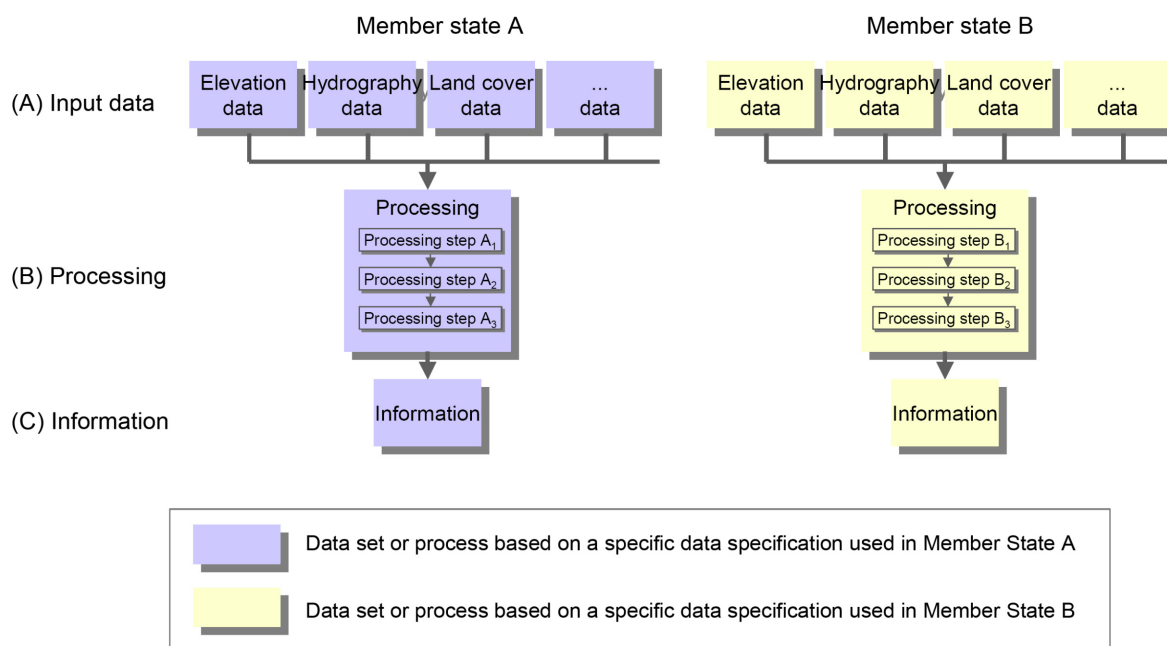


Рис. 5. Обычная ситуация: локальная (взаимонезависимая) обработка данных [INSPIRE-DSM, 2007]

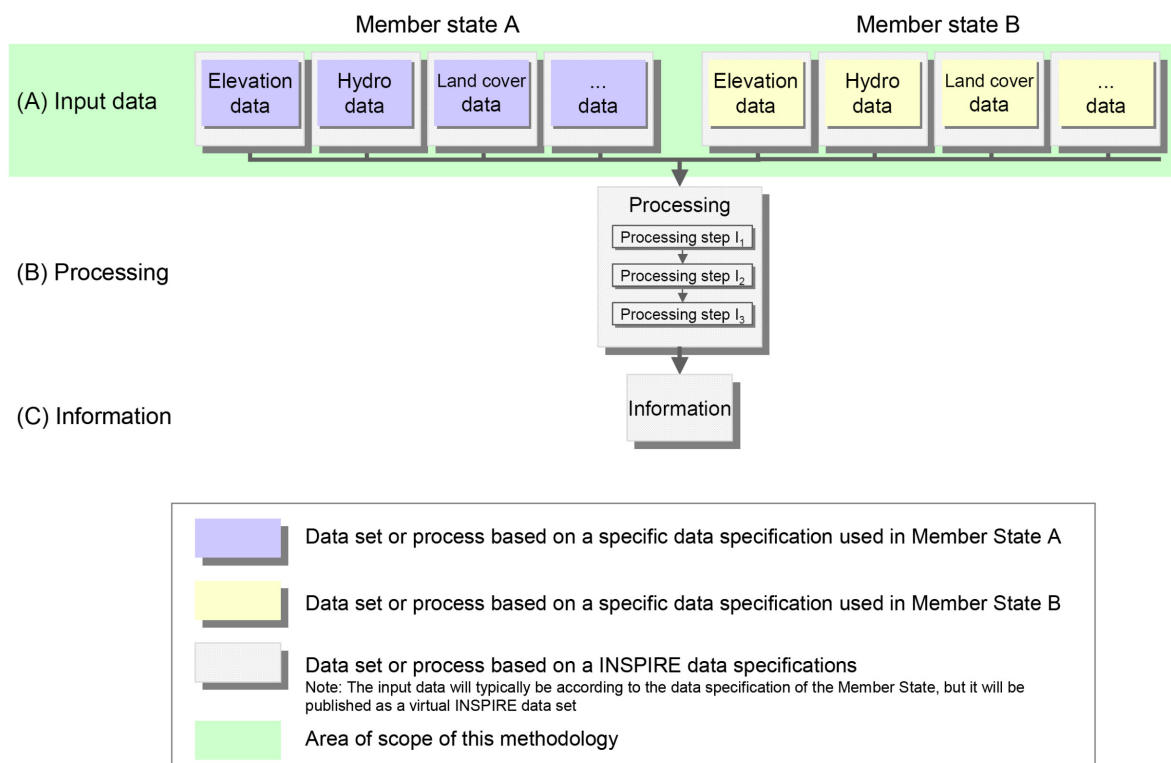


Рис. 6. Унифицированное представление данных, устраняющее взаимонезависимую обработку данных [INSPIRE-DSM, 2007]

INSPIRE - это инфраструктура для пространственной информации, которая создается государствами-членами, при которой:

- пространственные данные сохраняются, доступны и поддерживаются на наиболее приемлемом уровне;
- можно объединить пространственные данные из различных источников по всему сообществу последовательным образом и обмениваться ими между различными пользователями и приложениями;
- можно распределить пространственные данные, собранные одним органом государственной власти, между другими организациями;
- пространственные данные доступны для их широкого применения без дополнительных ограничений;
- легко обнаружить доступные пространственные данные, чтобы оценить их пригодность и узнать условия их использования.

Исходя из этих соображений, Директива фокусируется, в частности, на пяти ключевых областях:

- метаданные;
- возможность взаимодействия и согласования пространственных данных и услуг по отдельным темам (как описано в приложениях I, II, III части [INSPIRE Directive, 2007]);
- сетевые услуги и технологии;
- меры по обмену пространственными данными и услугами;
- меры по координации и мониторингу.

INSPIRE устанавливает правовую основу для создания и функционирования инфраструктуры для пространственной информации в Европе. Цель такой инфраструктуры – содействовать разработке политики по отношению к действиям, которые могут прямо или косвенно влиять на окружающую среду. Инфраструктура для пространственной информации включает:

- наборы метаданных о пространственных данных и услугах;
- сетевые услуги и технологий;
- соглашения о совместном использовании, доступе и использовании;
- координацию и мониторинг механизмов, процессов и процедур, устанавливаемых, эксплуатируемых или предоставляемых в соответствии с Директивой.

Напомним, что в Директиве [INSPIRE Directive, 2007] приняты следующие определения:

- "Метаданные" – информация, описывающая наборы пространственных данных и услуг, связанных с пространственными данными, и через которую возможно их обнаружение, систематизация и использование;
- "Пространственный набор данных" – идентифицируемое множество пространственных данных;
- "Пространственные данные" – любые данные, для которых имеется прямая или косвенная ссылка на конкретную точку земли или географического района;

- "Пространственный объект" – абстрактное представление некоторой сущности реального мира, связанной с конкретной точкой земли или географического района;

- "Сервисы для пространственных данных" – операции, которые могут быть выполнены над пространственными данными, содержащимися в наборах пространственных данных, или над соответствующими метаданными, путем вызова компьютерных приложений.

В принципе, каждый пространственный объект в пространственном наборе данных должен быть описан в спецификации, описывающей семантики данных и характеристики типов пространственных объектов в наборе данных [INSPIRE-TAO, 2007]. Типы обеспечивают классификацию пространственных объектов и определение среди прочей информации свойств, которые любой пространственный объект может иметь (тематические, пространственные, временные, функция покрытия и т.д.), а также известные ограничения (например, координатная реферативная система, которая может быть использована в наборах пространственных данных). Эта информация, в принципе, включена в схему приложения, используя язык концептуальной схемы, которая является частью спецификации данных.

В результате, спецификация данных обеспечивает необходимую информацию для интерпретации пространственных данных приложением.

Пространственный набор данных описывается набором метаданных, обеспечивающим поддержку обнаружения (а в определенной степени также оценку целесообразности использования) пространственных данных для конкретных целей.

Метаданные об сервисах предоставляют основную информацию о программах, предоставляющих услуги. Описание услуг включает в себя тип сервиса, описание операций и их параметров, а также информацию о географической информации, доступной через предлагаемые сервисы.

Формирование тематических слоев и их иерархической взаимосвязи в среде пространственных данных обеспечит взаимодействие между документами, содержащими пространственные данные, и архитектурами, ориентированными на услуги в глобальной сети. На основе выделенных из сетевых документов таксономий в среде пространственных данных могут быть сформированы соответствующие слои тематических карт. Сами пространственные данные, которые входят в соответствующий класс, становятся объектами соответствующего тематического слоя.

Напомним, что таксономия формируется на основе установления отношений между концептами и классами вида «группа объектов - объект». При этом контексты концептов, составляющих таксономию, определяют легенду карты, определяющей тематические применения пространственных данных. Благодаря объединению различных типов баз пространственных данных в таксономии, атрибуты объектов могут содержать гиперссылки на соответствующие документы и информационные системы глобальной сети.

Выводы

Трансдисциплинарная интеграция таксономии и тематической карты позволяет расширять и дополнять информационные описания понятий-объектов на базе сетевых информационных ресурсов, распределенных в интернет, а встроенная в таксономию поисковая машина - значительно расширить информацию об объектах. Такое сочетание позволяет создать единую информационно-аналитическую среду эксперта, которая перманентно пополняется территориально распределенными пользователями по различным направлениям исследований.

Предложенный подход особенно полезен при работе с пространственными данными. „Пространственные данные“ представляют собой данные с прямой или косвенной ссылкой на конкретную точку земной поверхности или географической области. Директива INSPIRE ЕС установила новый вид взаимодействия между странами-членами ЕС в области пространственной информации. Основной целью Директивы является интероперабельность ("Interoperability"), что означает возможность объединения пространственных наборов данных и возможность взаимодействия сервисов в автоматическом режиме. Таким образом увеличивается добавленная стоимость наборов данных и сервисов [INSPIRE Directive, 2007].

Использование трансдисциплинарного подхода к классификации, систематизации, использованию информационных ресурсов и интеграции распределенных информационных моделей и систем на основе семантических свойств пространственных данных дает возможность каждому пользователю обнаруживать принципиально новые, ранее неизвестные взаимосвязи, способствует смещению акцентов с пассивных методов поиска, ориентированных на передачу информации, к более широкому применению активных методов анализа проблем и поиска решений, сотрудничеству пользователей и разработчиков и экспертов.

Литература

- [Guarino, 1994] Guarino, N. The Ontological Level [Текст] / N. Guarino, R. Casati, N. Smith, G. White // Philosophy and the Cognitive Sciences. – Vienna : Holder-Pichler-Tempsky, 1994. – p. 443-456.
- [INSPIRE Directive, 2007] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) [Электронный ресурс]. Режим доступа: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>
- [INSPIRE-DSM, 2007] INSPIRE Data Specifications: Methodology for the Development of Data Specifications [Электронный ресурс]. Режим доступа: http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/inspireDataspesD2_6v2.0.pdf
- [INSPIRE-TAO, 2007] INSPIRE Technical Architecture Overview. [Электронный ресурс]. Режим доступа: http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/INSPIRETechnicalArchitectureOverview_v1.2.pdf
- [Report, 2012] World Economic Forum. Insight Report [Электронный ресурс] / The Global Information Technology Report 2012. Living in a Hyperconnected World. Режим доступа: http://www3.weforum.org/docs/Global_IT_Report_2012.pdf

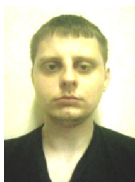
- [Барендрегт, 1985] Барендрегт Х. Лямбда-исчисление. Его синтаксис и семантика: Пер. с англ. — М.: Мир, 1985. — 606 с.
- [Белов, 2005] Белов В. С. Информационно-аналитические системы. Основы проектирования и применения: учебное пособие, руководство, практикум / Московский государственный университет экономики, статистики и информатики. — М., 2005. — 111 с.
- [Буч, 1992] Буч Г. Объектно-ориентированное проектирование с примерами применения: Пер. с англ. — М.: Конкорд, 1992. — 519 с.
- [Величко, 2009] Величко В. Автоматизированное создание тезауруса терминов предметной области для локальных поисковых систем / В. Величко, П. Волошин, С. Свитла // "Knowledge – Dialogue – Solution" International Book Series "INFORMATION SCIENCE & COMPUTING", Number 15. – FOI ITHEA Sofia, Bulgaria. - 2009. – pp.24-31.
- [Величко, 2014] Величко В. Ю. Комплексные инструментальные средства инженерии онтологий / В. Ю. Величко, К. С. Малахов, В. В. Семенов, А. Е. Стрижак // International Journal «Information Models and Analyses», 2014. – Volume 3. – Number 4. – P. 336-361.
- [Гаврилова, 2000] Гаврилова, Т. А. Базы знаний интеллектуальных систем [Текст] / Т. А. Гаврилова, В. Ф. Хорошевский. — СПб.: Питер, 2000. — 384 с.
- [Гладун, 1994] Гладун, В. П. Процессы формирования новых знаний [Текст] / В. П. Гладун. — София : СД «Педагог 6», 1994. — 192 с.
- [Глушков, 1982] Глушков, В. М. Основы безбумажной информатики / В. М. Глушков. — М. : Наука, 1982. — 552 с.
- [Загоруйко, 2008] Загоруйко, Ю. А. Автоматизация сбора онтологической информации об интернет-ресурсах для портала научных знаний [Текст] / Ю. А. Загоруйко // Известия Томского политехнического университета. — Т. 312. — № 5. Управление, вычислительная техника и информатика. — 2008. — с. 114-119.
- [Клини, 1957] Клини, С.К. Введение в метаматематику [Текст] / С. К. Клини. — М. : Иностранная литература, 1957. — 526 с.
- [Коршунова, 2009] Коршунова, С. О. Роль тезаурусного моделирования в организации терминополья «ТЕХТ-ТЕКСТ» [Электронный ресурс] / С. О. Коршунова // Вестник Иркутского государственного лингвистического университета, 2009. — № 1. — Режим доступа: <http://cyberleninka.ru/article/n/rol-tezaurusnogo-modelirovaniya-v-organizatsii-terminopolya-text-tekst>.
- [Малишевский, 1998] Малишевский, А.В. Качественные модели в теории сложных систем [Текст] / А. В. Малишевский. — М. : Наука. Физматлит, 1998. — 528 с.
- [Палагин, 2012] Палагин, А. В. Онтологические методы и средства обработки предметных знаний [Монография] / А. В. Палагин, С. Л. Крытый, Н. Г. Петренко. — Луганск: изд-во ВЛУ им. В. Даля, 2012. — 323 с.
- [Стрижак, 2014] Стрижак А. Е. Операциональные характеристики онтологий / А. Е. Стрижак // Открытые информационные и компьютерные интегрированные технологии, 2014. — № 66. — с. 185–193.
- [Стрижак, 2015] Стрижак О.Е. Трансдисциплінарна інтеграція інформаційних ресурсів // Автореферат дисертації на здобуття наукового ступеня доктора технічних наук. -К.- Інститут телекомунікацій і глобального інформаційного простору НАН України.-2015.- 42с.
- [Шаталкин, 2012] Шаталкин, А.И. Таксономия. Основания, принципы и правила [Текст] / А. И. Шаталкин. — М. : Товарищество научных изданий КМК, 2012. — 600 с.

Информация об авторах



Величко Виталий Юрьевич – к.т.н., старший научный сотрудник, Институт кибернетики им. В. М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: velychko@aduis.com.ua

Основные области научных исследований: индуктивный логический вывод, обработка естественно-языковых текстов



Приходнюк Виталий Валериевич – аспирант, Институт телекоммуникаций и глобального информационного пространства НАН Украины, Киев-186, 03186, Чоколовский бульвар, 13; e-mail: vitalik1700@yandex.ru

Основные области научных исследований: геоинформационные системы, онтологический инжиниринг



Стрижак Александр Евгеньевич – д.т.н., ведущий научный сотрудник, Институт телекоммуникаций и глобального информационного пространства НАН Украины, Киев-186, 03186, Чоколовский бульвар, 13; e-mail: sae953@gmail.com

Основные области научных исследований: корпоративные интеллектуальные системы, поддержка принятия решений



Крассимир Марков – доцент, д-р, Институт математики и информатики БАН, София, Болгария; e-mail: markov@foibg.com

Основные области научных исследований: общие теоретические проблемы информатики, многомерные информационные пространства и базы.



Крассимира Иванова – доцент, д-р, зав. каф. Информационные технологии, Университет телекоммуникаций и почт, София, Болгария; e-mail: krazy78@mail.bg

Основные области научных исследований: информационные технологии, многомерные многослойные информационные структуры.



Стефан Карастанев – ассистент, Институт механики БАН, София, Болгария; e-mail: stefan@imbm.bas.bg

Основные области научных исследований: информационные системы и базы данных.