УДК:519.766.23:005.94.378.141

ТЕОРЕТИКО-МНОЖЕСТВЕННАЯ МОДЕЛЬ УПРАВЛЕНИЯ ЗНАНИЯМИ

БАБАК В.Ф., КАСЫМАЛИЕВА А.Т., ТОРОБЕКОВ Б.Т.

КГТУ им. И. Раззакова

izvestiya@ktu.aknet.kg

Краткая аннотация. В третьей, заключительной статье из цикла компьютерно-ориентированных принципов моделирования инновационного ВУЗа описана современная точка зрения на информационный менеджмент и управление знаниями. В основу управления знаниями положена теория семантических сетей.

Современная парадигма *управления* информационными системами и системами управления знаниями (СУЗ) включает компоненты:

- сбора и накопления (ввод данных и знаний),
- хранение,
- преобразование,
- передача данных и знаний по требованию в нужное время и место,
- фрагментирование под задачи анализа и принятия решений,
- навигацию и поиск,
- документирование.

Среди перечисленных компонент СУЗ критичной и наименее исследованной является система навигации и поиска знаний, соотносимых с каждым классификатором уровня бизнес-модели И-ВУЗа, поскольку эта система должна обеспечить доступ не только к одной из перечисленных выше программных систем и технологий, но и к нужному, конкретному разделу знаний. Среди небольшого количества таких систем наиболее известными являются «объектная модель документов ***DOM*** (*Document Object Model* — «объектная модель документа»)**»** и «*навигационные карты знаний*», концептуальную основу которых составляют семантические сети.

Модель DOM не налагает ограничений на структуру документа. Любой документ известной структуры с помощью DOM может быть представлен в виде дерева узлов, каждый узел которого представляет собой элемент, атрибут, текстовый, графический или любой другой объект. Узлы связаны между собой отношениями "родительский-дочерний".

Однако значение *навигационных карт знаний* в большей степени носит *методологический характер*, заключающийся в том, что на естественном языке, с помощью словарей понятий и фиксированного набора отношений, можно конструировать сложные системы (сети) представления знаний, позволяющие выполнять навигацию и отслеживать полноту связей между наборами классификационных деревьев и множеством компонент знаний.

На схеме (рис.1) приведен фрагмент *авторского варианта семантической сети* (карты знаний), имеющий отношение к рассматриваемой проблеме – моделированию ресурсного и управленческого аспекта *пространства знаний* современного И-ВУЗа.

Язык моделирования включает следующие элементы сети:

- в квадратах обозначены имена классов (категорий) элементов, входящих в семантическую сеть;
- в необрамленных квадратами текстах представлены типы реализации классов - типы и объекты хранения экземпляров классов, называемые спецификаторами.
- именованными стрелками помечены типы отношений «класс – класс» и «класс - экземпляр класса».

Отношения между классами элементов сети, а также классами и экземплярами классов на языке теории отношений могут быть представлены формализмами, имеющими глобальный характер:

$$R \subset X \times Y, \text{ где}$$

X, Y– формальные образы классов, реализуемые функциями интерпретации *и* $\varphi_1 : X \to$ <*Имя класса*> и $\varphi_2 : Y \to$ <*Имя класса*>.

Применив функции интерпретации к классу **Классификатор,** связанному отношением *имеет с* классами **Имя** и **ID классификатор** , получим

$R_{\text{имеет}} \subset$ <*Классификатор*> $\times$ <*Имя*> и

$R_{\text{имеет}} \subset$ <*Классификатор*> $\times$ < *ID классификатор* >

Пусть далее отношение имеет смысл $R \subset X \times Y$, где X – формальные образы семейства классов, Y – формальный образ семейства экземпляров класса, реализуемые функциями интерпретации *и* $\varphi_1 : X \to$ <*Имя класса*> и $\varphi_2 : Y \to$ <*Экземпляр класса*>, тогда

1) выражение $\forall x \forall y\ (\boldsymbol{P}(x,y)) : x \in X\ \&\ y \in Y$ *и* $\boldsymbol{P}$ = <*есть*> отвечает *условию полноты* отношения R между классами и экземплярами классов;

2) выражение $\forall x\ \exists y\ (\boldsymbol{P}(x,y)) :$ $x \in X\ \&\ y \in Y$ *и* $\boldsymbol{P}$ = <*есть*> отвечает *условиям существования* экземпляров класса $y \in Y$ для класса X;

3) выражение $\exists x\ \exists y\ (\boldsymbol{P}(x,y)) :$ $x \in X\ \&\ y \in Y$ *и* $\boldsymbol{P}$ = <*есть*> отвечает условию *полного соответствия* отношений между классами и экземплярами классов.

Аналогичные формализмы могут быть введены для других имен отношений R: <*класс – класс*> и< *класс – экземпляр класса*>, как то <*входит в*>, <*содержит*>, <*имеет*> и пр.

Отметим, что приведенные выше условия 1) -3) для типов отношений, где классы элементов сети соотносятся с их экземплярами, имеющими тип словаря, справочника, массива или базы данных, являются необходимыми, но далеко недостаточными. Достаточность этих отношений будет показана после формулировки теорем, приведенных далее.

На уровне программных реализаций отмеченные отношения представляются матрицами соответствий, а функции управления отношениями (визуализация, навигация, выбор, отношение следования) задаются в табличной форме, путем привязки этих отношений к соответствующим классификаторам.

По отношению к *Карте знаний* можно определить такие же отношения и условия существования между *категориями (классами) знаний и спецификаторами (экземплярами классов),* входящими в состав этих категорий. Применительно к схеме семантической сети, приведенной на рис.1, к *разряду категорий знаний* относятся вершины с именами *Атрибуты, Файлы, Данные, Связи и* пр., а к разряду спецификаторов, связанные с категориями – *экземпляры (реализации) категорий.*

Существует несколько типов отношений между экземплярами классификаторов и спецификаторов. Рассмотрим два из них.

Отношение «дерево-дерево». Пусть существует некоторый классификатор $D\ (V,\ E)$, имеющий структуру размеченного дерева и принадлежащий лесу деревьев $\boldsymbol{D}$, моделирующему некоторую организацию (И-ВУЗ). Каждый узел $v_i\ \in V$ этого классификатора содержит *имя* узла и *ID-идентификатор*, т.е. специфицируется упорядоченной двойкой («*ID*», *«имя»).* Пусть далее каждому дереву D из леса деревьев $\boldsymbol{D}$ поставлено в соответствие некоторое дерево $A\ \in \boldsymbol{Kw},$ изоморфное дереву $D \in \boldsymbol{D}$, тогда между деревьями D и A существует отношение

$$R\ \subset \times\{D,\ A\}, \qquad (1)$$

удовлетворяющее условию

$\forall D(V,E) \in \boldsymbol{D}\ \exists A(V,E) \in \boldsymbol{Kw}\!: P(D,\ A)\ \&\ P \Leftrightarrow$ *«содержит»* $\vee$ *«имеет».*

ABBYY PDF Transformer 2.0
Click here to buy
www.ABBYY.com

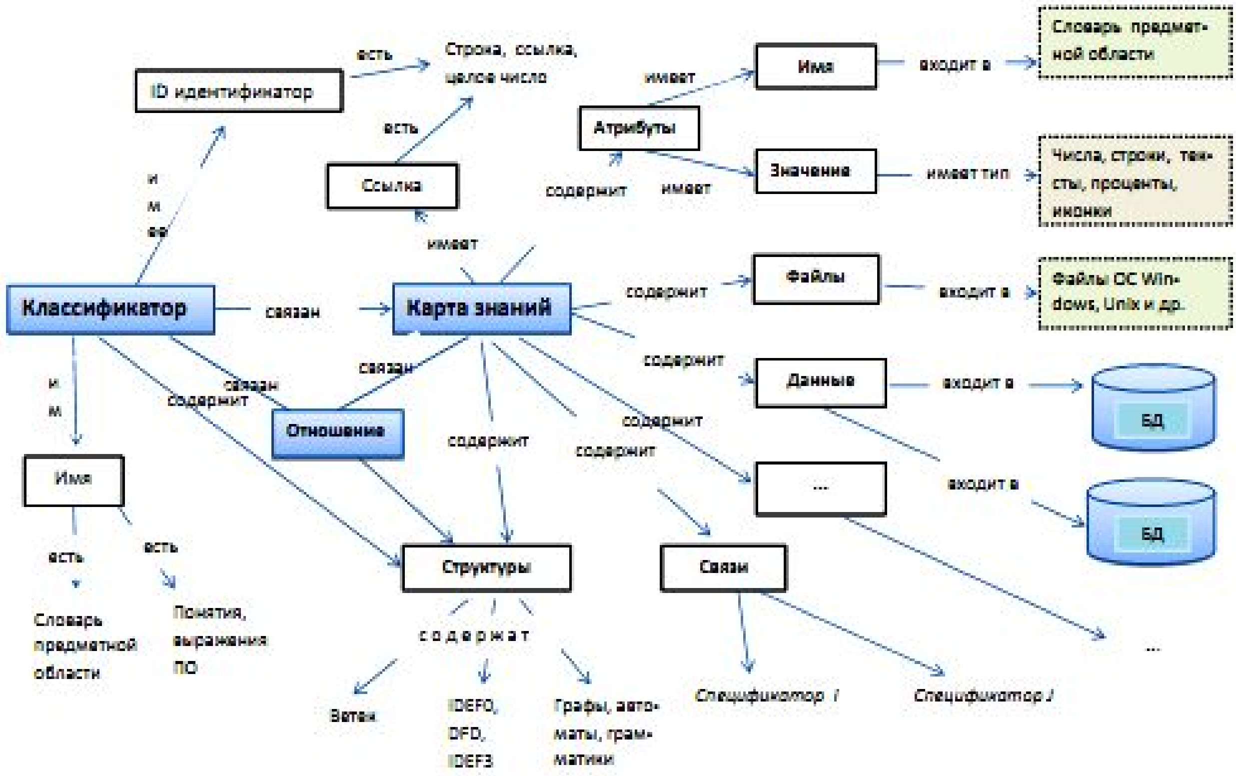

Рис.1.Семантическая сеть в форме «Карты знаний» модели И-ВУЗа (фрагмент)

Отношение «дерево-множество». Пусть существует некоторый классификатор $D(V, E)$, имеющий структуру размеченного дерева и принадлежащий лесу деревьев $\boldsymbol{D}$, моделирующему некоторую организацию (и, в частности, И-ВУЗ). Каждому D соответствует некоторое множество A, такое, что каждый элемент этого множества $a \in A$ специфицируется набором из m (множеством) элементов $\{a_1, a_2, \ldots, a_i, \ldots, a_m\}$; в общем случае каждый элемент $a \in A$ имеет различное число m принадлежащих ему элементов. Примем, что имеет место некоторое отображение D на (или в) A

$$F: \ D \rightarrow A, \qquad (2)$$

такое, что

$$\forall D(V,E) \in \boldsymbol{D} \ \exists A \in \boldsymbol{Kw}: P(D, A) \ \& \ P \Leftrightarrow «содержит» \vee «имеет»$$

Введем следующие определения [1,2,3,4]:

<u>*Определение 1.*</u> **Классом эквивалентности** $C(a)$ элемента a назовем подмножество элементов, эквивалентных a. Из этого определения следует, что, если $b \in C(a)$, то $C(a) = C(b)$. Множество всех классов эквивалентности будем обозначать $X/\sim$, а для класса эквивалентности элемента a, - использовать обозначение $a/\sim$. Множество классов эквивалентности по отношению $\sim$ определяет и задает разбиение множества X.

Определение 2. Множество классов эквивалентности, отвечающее отношению эквивалентности **~** , и обозначаемое символом X/**~** будем называть *фактор-множеством* относительно **~**.

Определение 3. Пусть дерево $D(V,E) \in D$ специфицируется семейством множеств $A = \{a\}$ из категории классификаторов **Атрибуты**, т.е. $A \in \boldsymbol{Kw}$ и пусть каждый набор (множество) атрибутов $a \in A$, специфицирующих некоторый узел v дерева D из леса деревьев $\boldsymbol{D}$, задан множеством элементов $a_i \in a$, называемых атрибутами, т.е. $a = \{a_1, a_2, ..., a_i , ..., a_m\}$, **,** где, в общем случае, для каждого узла v дерева D количество m атрибутов a_i различно. Назовем каждый атрибут a_i из множества $a \in A$ атрибутов, специфицирующих узел v дерева D, *специфицирующим атрибутом (СА)*, а множество атрибутов $a \in A$, принадлежащих категории A **Атрибуты** – *специфицирующим множеством атрибутов* (*СМА)* или *атрибутизированным спецификатором знаний (АСЗ).*

Определение 4. Назовем поддеревом D_i часть дерева D, которая может быть представлена в виде отдельного дерева. Любой узел дерева Di вместе со всеми его узлами-потомками является поддеревом дерева D. Для любого узла поддерева либо должен быть путь в корневой узел этого поддерева, либо сам узел должен являться корневым. Тогда каждое поддерево D_i имеет максимальный элемент $max(D)$, рассматриваемый как само дерево D и набор минимальных элементов $min(D)$ - концевые узлы (листья) дерева D.

Атрибутизация классификаторов. Рассмотрим категорию спецификаторов **Атрибуты** и варианты того, как спецификаторы этого класса соотносятся с элементами классификационных деревьев D ∈ D.

Предложение 1. Существуют структуры классификаторов $D \in \boldsymbol{D}$ из леса классификаторов $\boldsymbol{D}$, которые могут быть специфицированы *единым набором атрибутов* $a \in A$ *СМА,* необходимым и достаточным, чтобы специфицировать каждый узел v дерева D_i . Это предложение поддерживается следующей теоремой.

Теорема 1. Пусть существуют классификаторы D из леса семантически окрашенных классификаторов $\boldsymbol{D}$, где $D \in \boldsymbol{D}$, для которых все узлы $v \in V$ (вершины) дерева $D(V,E)$ специфицируются одним и тем же набором атрибутов $a \in A$, определенным на классе эквивалентости a /**~** , тогда существует отображение

$$F : D \rightarrow a/\sim \qquad (3)$$

такое, которое удовлетворяет условию

$\forall v \in D\ \exists a/\sim\ : P(v, a/\sim)$, где $P \Leftrightarrow$ *«имеет»* $\vee$ *«содержит».*

Доказательство. Пусть $D\ (V,E)$ –семантически окрашенное дерево леса деревьев $\boldsymbol{D}$ и $D \in \boldsymbol{D}$, соответствующее некоторому базовому классификатору с числом узлов $n = |\ v\ |$ Переименуем дерево D в дерево $H =(V, E)$. Поставим в соответствие дереву H граф $K =\{A,E\}$, каждый узел a которого представлен множеством m атрибутов $\{a_i : i\ = 1..m\}$, где – i множество целых чисел. В общем случае для каждого узла a графа K число атрибутов m различно.

ABBYY PDF Transformer 2.0
Click here to buy
www.ABBYY.com

Пусть граф $H = (V, E)$ *изоморфен* графу $K = (A,E)$, так что существует *биекция* f из множества вершин графа H в множество вершин графа Kn, обладающая следующим свойством: если в графе H есть ребро из вершины v_i в вершину v_{i+1}, то в графе Kn должно быть ребро из вершины $f\,(a_i)$ в вершину $f\,(a_{i+1})$ и наоборот – если в графе Kn есть ребро из вершины a_i в вершину a_{i+1}, то в графе H должно быть ребро из вершины $f^{-1}\,(v_i)$ в вершину $f^{-1}\,(v_{i+1})$.

Определим некоторое глобальное множество $\boldsymbol{\Lambda}$, составляющее словарь специфицирующих атрибутов, где каждый набор атрибутов $a \subset \boldsymbol{\Lambda}$ и каждый атрибут a_i набора a – есть упорядоченная двойка <(*имя атрибута= значение*)>.

Пусть далее граф $H = (V, E)$ - есть дерево классификаторов, где каждый его узел v – имя некоторого классификатора, входящее в иерархию классификаторов с базовым классификатором v_0, представляющим корневой узел графа; а каждый узел a графа $K = (A,E)$, изоморфного графу $H = (V, E)$, специфицируется множеством $a = \{a_1, a_2, \ldots, a_i, \ldots, a_m\}$, составленным из набора специфицирующих этот узел имен атрибутов a_i, принадлежащих множеству элементов словаря $\boldsymbol{\Lambda}$, и $a_i \in \boldsymbol{\Lambda}$, тогда имеет место биективное отображение $f: v \to a$ классификатора v на множество a или $a = f\,(v)$.

Пусть порядок (число узлов) графа $|H|$ равен n, т.е. $n = |H|$. Выполним над узлами a_i графа K, где $a_i \in K$, операцию объединения таким образом, что $a' = a_{i1} \cup a_{i2} \cup, \ldots, \cup a_{in}$ образует множество a'. Проведем факторизацию множества a', удалив (возможно и переименовав) в a' одноименные по смыслу атрибуты, оставив их по одному, тогда мощность множества атрибутов $|a|$ определится на классе эквивалентности $a = a'/\boldsymbol{\sim}$.

Сюръективное отображение $F\!: H \to a$, удовлетворяющее условию

$$\forall v \in H\ \exists a = a'/\boldsymbol{\sim}:\ P(v, a), \text{ где } P \Leftrightarrow «имеет» \vee «содержит»,$$

вполне обеспечивает атрибутизацию всех узлов v дерева G. На схемном уровне это представляется отображением, приведенном на рис. 2.

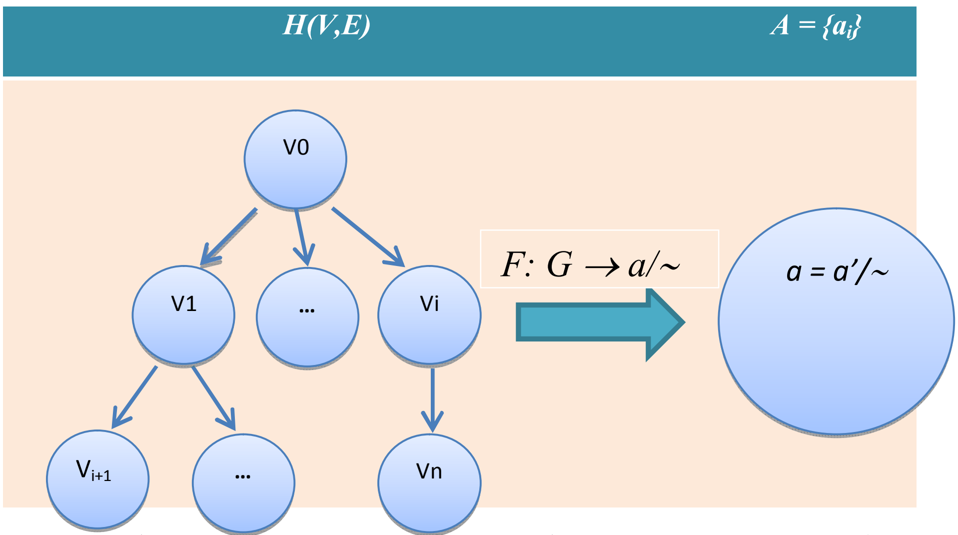

Рис.2. Отображение $F\!: G \to a = a'/\sim$, атрибутизирующее все узлы v дерева $G(V,E)$

Теорема 2. Существуют классификаторы D из леса семантически окрашенных классификаторов $\boldsymbol{D}$, где $D \in \boldsymbol{D}$, для которых каждое поддерево $D_i(V_i,E_i)$ (вершины) дерева $D(V,E)$ специфицируются одним и тем же набором атрибутов A_i, определенным на классе эквивалентности $A_i/\sim$.

Доказательство. Доказательство теоремы 2 аналогично доказательству теоремы 1 с отличием того, что факторизация спецификаторов $A_i = A'_i/\sim$ выполняется на подмножествах атрибутов A'_i, соотносимых с каждым узлом поддерева D_i . Таким образом, согласно теореме 1 существует сюръективное отображение F_i: $G_i \to A_i$, удовлетворяющее условию

$\forall v_i \in G_i\ \exists A_i = A_i'/\sim : P(v_i, A_i)$, где $P \Leftrightarrow$ *«имеет»*,

которое вполне обеспечивает атрибутизацию всех узлов v_i поддерева G_i.

На схемном уровне это представляется отображением, приведенном на рис. 3.

Теорема 3. Существуют классификаторы D из леса семантически окрашенных классификаторов $\boldsymbol{D}$, где $D \in \boldsymbol{D}$, для которых каждый узел v_i (вершина) дерева $D(V,E)$ специфицируются уникальным набором атрибутов a_i, специфицирующих только и только этот узел.

Доказательство этой теоремы очевидно, а биективное отображение F_i: $v_i \to a_i$, удовлетворяет условию,

$\forall v_i \in G\ \exists a_i \in \boldsymbol{\Lambda}: P(v_i, a_i)$, где $P \Leftrightarrow$ *«имеет»*,

что иллюстрируется схемой, приведенной на рис.3.

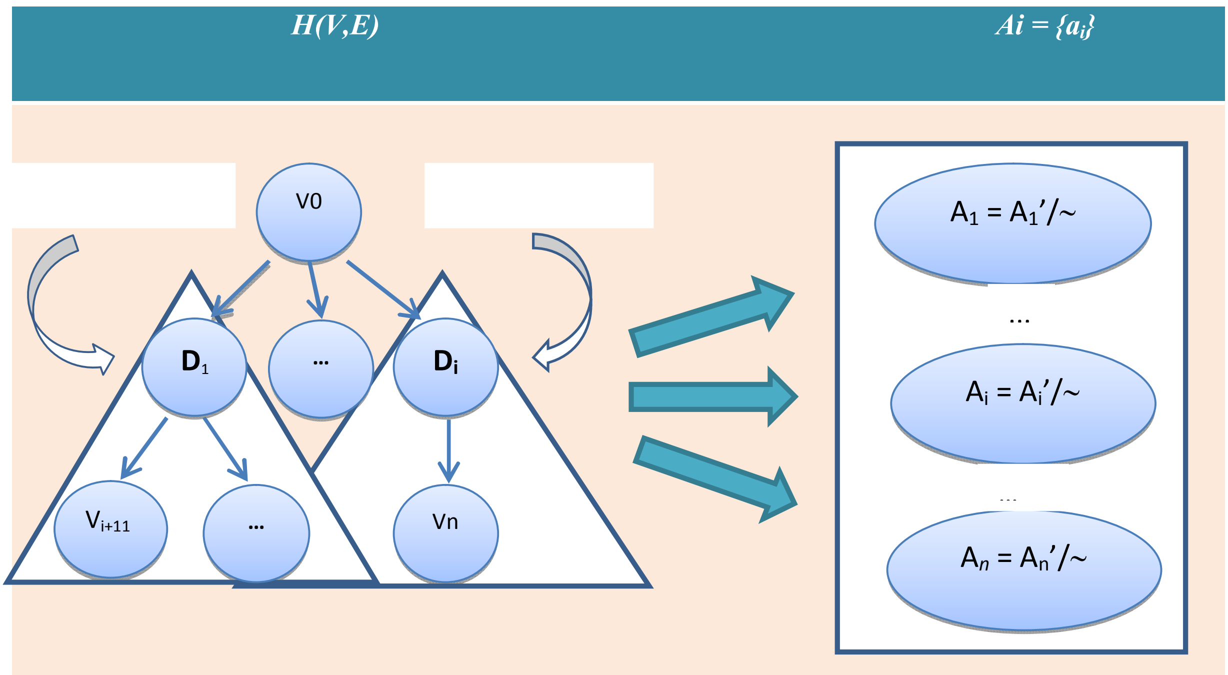

Рис.3. Отображение F: H $\rightarrow$ A =A'/~, атрибутизирующее все узлы v дерева H(V,E).

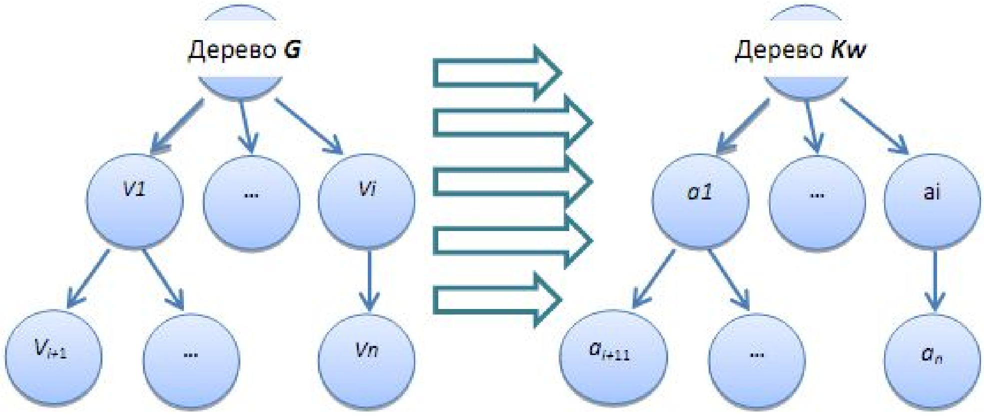

Рис.4. Отображение F_i: $v_i \rightarrow a_i$, атрибутизирующее все узлы v дерева G(V,E)

Связывание классификаторов с файлами.

Для создания ресурсной части знаний и управления знаниями одной атрибутизации недостаточно. Необходимы более мощные средства представления знаний, и в качестве таковых могут использоваться файловые, Web-представления, базы данных и знаний, а также техники управления этими знаниями.

В качестве базовой модели представления и управления знаниями также примем формализм в формате парных изоморфных деревьев (см. рис. 5).

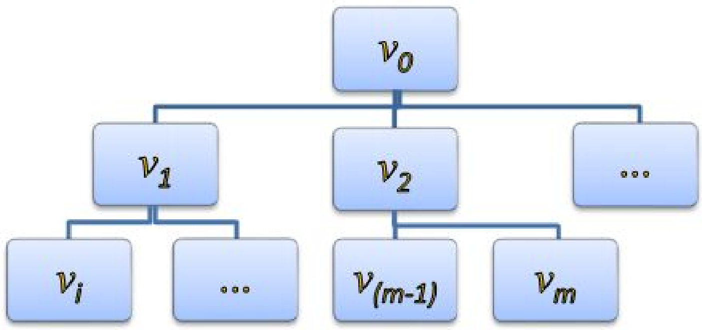

Рис.5а. Дерево G(V,E), идентифицирующее узлы классификатора

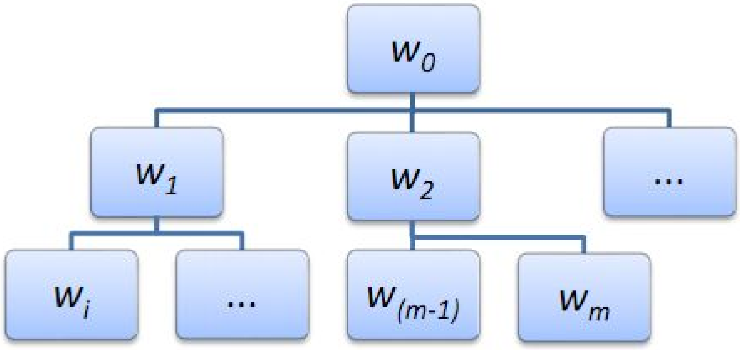

Рис.5b. Дерево Kw(W,E), специфицирующее категорию ***Файлы.***

Деревья $G(V,E)$ и $Kw(W,E)$ входят в леса деревьев ***G*** *и* ***Kw*** соответственно, так что – $G(V,E) \in$ ***G*** и $Kw(W,E) \in$ ***Kw.*** Каждый узел v дерева $G(V,E)$ представлен двойкой *<индекс, идентификатор>*; а каждый узел w дерева $Kw(W,E)$ – набором файлов $w = \{w_1, w_2, \ldots, w_i, \ldots, w_n\}$, где каждый файл w_i описывается также двойкой *<путь доступа к файлу, дескриптор>*, а число файлов n в разных наборах w, в общем случае, различно. Не исключается и ситуация, когда набор файлов отдельного узла w дерева $Kw(W,E)$ пуст ($\varnothing$).

Связи между узлами $v \in V$ и $w \in W$ деревьев $G(V,E)$ и $Kw(W,E)$ удобно задавать в виде отношения

$$R_{G,Kw} \subset G(V,E) \times Kw(W,E), \qquad (4)$$

реализуемого на программном уровне матрицей соответствия

$$F_{G,Kw} : G(V,E) \rightarrow Kw(W,E). \qquad (5)$$

В общем случае функция (3.5) является частичной, т.е не каждый узел $v \in V$ дерева $G(V,E)$ имеет специфицирующий его набор файлов ($w_i = \varnothing$).

Что касается избирательного специфицирования узлов деревьев $G(V,E)$ спецификаторами $Kw(W,E)$ из категории **Файлы**:

- одним набором w_0 на все дерево $G(V,E)$ из леса деревьев ***D:*** $G(V,E) \in$ ***D;***
- по одному набору w_i на некоторые поддеревья $G_i\,(V,E)$ дерева $G(V,E)$ из леса деревьев ***D***: $G_i\,(V,E) \in G\,(V,E)$;
- по одному набору w на каждый узел $v \in G(V,E)$ дерева $G\,(V,E)$ из леса деревьев ***D***,

то имеет место следующее предложение.

Предложение 1. Теоремы существования 1 - 3, доказанные по отношению к спецификации деревьев $G(V,E) \in$ ***D*** спецификаторами $K(A,E)$ из категории **Атрибуты**, справедливы и при спецификации деревьев $G(V,E) \in$ ***D*** спецификаторами $Kw(W,E)$ из категории **Файлы.**

Доказательство теорем очевидно, если заменить обозначения деревьев $K(A,E)$ на деревья $Kw(W,E)$.

Связывание классификаторов с web-документами. Эта связь реализуется на уровне гиперссылок на нужные сайты со стороны узлов различного уровня представления классификационных деревьев. Здесь также справедливы теоремы 1 - 3 с заменой экземпляров спецификаторов из категории **Файлы** на категорию **Web- документы**.

Категория **Web- документы** относится к спецификаторам внешней формы представления ресурсов знаний по отношению к знаниям организации. Управление этими ресурсами реализуется глобальными механизмами навигации и поиска, развиваемыми в теории, получившей название *Всемирной семантической паутины,* целью которой является представление информации (знаний) в виде, пригодном для машинной обработки и, в то же время, использующей не только механизмы гиперссылок, но и семантику содержания.

Введение в *Пространство знаний* категории **web-документы** позволяет представлять и обрабатывать внешние ресурсы знаний обеспечивая предметно-ориентированное сужение этого пространства, привязывая его к тому или иному уровню классификационных деревьев. Особенно значим такой подход для концепции воспроизводства знаний, в условиях *глобализации и унификации образовательного контента* преподаваемых дисциплин.

Литература

1. Бабак В.Ф., Касымалиева А.Т., Торобеков Б.Т. Теоретико-множественная модель И-ВУЗа.-- Публикация в этом же сборнике, 2013 г**.**
2. Месарович М., Такахара Я. Общая теория систем: математичеcке основы.- Перев.с англ. Под редакцией акад. Емельянова С.В., Инд-во «Мир», 1978 г., 311 с.
3. Харари Ф. Теория графов. — М.: Мир. — 1973. — 300 с.

Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ = Introduction to Algorithms / Под ред. И. В. Красикова. — 2-е изд. — М.: Вильямс, 2005. — 1296 с.