УДК 004.827, 519.584

Е.П. Ильина, О.А. Слабоспицкая

ФОРМЫ, МЕТРИКИ И СВОЙСТВА ОТНОШЕНИЯ СХОДСТВА МЕЖДУ КОНЦЕПТАМИ В ОНТОЛОГИЯХ ЭКСПЕРТНЫХ ТОЧЕК ЗРЕНИЯ

Рассмотрены формализация и метризация отношения сходства между концептами, которые служат в онтологиях концептуально различных точек зрения представлением элементов постановки экспертной проблемы (оцениваемого показателя, контекста решения и эталонов верификации). Предложен ряд форм отношения сходства и функции оценки для них. Исследованы свойства отношений и оценок. Показана конструктивность предложенного формализма для создания процедур сравнения концептуально различных экспертных точек зрения на проблему.

Постановка задачи. Одним из широко применяемых подходов к принятию решений является экспертный, привлекательность которого обусловлена возможностью непосредственного использования профессионального опыта специалистов [1] при решении задачи, поставленной инициатором экспертизы. На практике экспертная методология является незаменимой для решения проблем, лежащих на стыке нескольких областей знания, в условиях слабоформализованного знания о предметной области (ПрО), в которой широко используются эмпирика и эвристические методы, а также при изменяющихся знаниях и представлениях по поводу объекта экспертизы.

Однако имеется серьезная проблема в решении задач экспертного оценивания, касающихся такой целевой характеристики (ЦХ) объекта экспертизы, которая не относится к объективно существующим, а принадлежит к социальным конструктам субъективной оценочной природы [2]. Обобщение индивидуальных экспертных оценок в этом случае становится более чем затруднительным, поскольку постулируемая в имеющихся методах [3] однородность индивидуальных оценок, как и их сопоставимость в более широком понимании, требует, чтобы с онтологической точки зрения оценки были экземплярами одного и того же концепта онтологии ПрО экспертизы.

Пусть имеются две точки зрения VP_i, VP_j на ПрО экспертизы, каждая из которых имеет свое онтологическое представление ПрО $O(VP_i)$, $O(VP_j)$, соответствующее профессиональной деятельности группы специалистов, разделяющих эту точку зрения.

Пусть оцениваемая характеристика отображается концептом c, принадлежащим обеим онтологиям с общим вектором его свойств $\{c_i\}$. Пусть также имеется множество $X(VP_i)$ концептов, связанных в рамках $O(VP_i)$ с концептом c связями φ, которые определяют значения $\{c_i\}$ на основе значений X, и множество $Y(VP_j)$ концептов, выполняющих ту же роль в $O(VP_j)$. Тогда экспертные оценки $c^*(VP_i)$ и $c^*(VP_j)$ для одного и того же объекта экспертизы в одних и тех же условиях среды (часть которых отражается в элементах X, а часть – в элементах Y) будут систематически, а не случайно, различными и, кроме того, иметь разную тенденцию реагирования на изменения условий задачи. Это сделает их несопоставимыми. Какая бы процедура обобщения (статистического или метрического) ни применялась, при отсутствии процедур установления онтологических эквивалентностей она не предоставит никаких значимых критериев приемлемости полученной обобщенной оценки.

При реализации информационно-аналитической деятельности в различных сферах государственного управления [4], а также в пределах циклов программно-целевого планирования [5] экспертное оценивание приобретает характер регулярно выполняемой процедуры, имеющей дело с объектами экспертизы одних и тех же классов. Круг профессиональных и ведомственных групп, точки зрения которых на объект экспертизы могут представлять интерес, является в этом случае достаточно стабильным.

В описанных условиях решение проблемы обобщения индивидуальных мнений можно основывать на априорном анализе знаний о собственной деятельности, прямо или опосредованно касающейся объектов экспертируемых классов, которыми руководствуются носители актуальных точек зрения. Ниже такие системы знаний будут рассматриваться как онтологии экспертных точек зрения, основанные на концептуальной модели специального вида.

Экспертизы такого рода должны сопровождаться аналитиком. Онтологический анализ предоставит ему возможность судить о гипотетических рисках применения той или иной постановки и стратегии решения экспертных задач в условиях привлечения к экспертизе представителей выбранных точек зрения.

Средством получения сопоставимых экспертных оценок в предлагаемом подходе является формирование единой концептуальной модели решаемой проблемы, которое может осуществляться в рамках различных стратегий аналитического сопровождения процесса экспертизы. В состав этой модели входят:

- проблемный вопрос, определяющий оцениваемые объекты и тип ЦХ оценивания;

- модель ЦХ, задающая ее трактовку в терминах других концептов ПрО;

- контекст решения проблемы, являющийся перечислением концептов ПрО, информа-ция из которых должна приниматься во внимание при решении;

- поле верификации, содержащее перечисление концептов ПрО, сопоставление с кото-рыми позволяет судить о допустимости полученного результата, предоставляя тем самым его обоснование.

Наиболее предпочтительной стратегией (обозначим ее S_1) является формирование аналитиком предоставляемой экспертам постановки проблемы, при использовании которой снижаются гипотетические риски несопоставимости экспертных оценок. Для этого аналитик должен конструктивно реализовать следующие принципы формирования постановки на основе онтологий ПрО, соответствующих разным точкам зрения и отражающих опыт решения проблем, аналогичных данной в аспекте проблемного вопроса.

1. Концепты, используемые в проблемном вопросе, должны непосредственно присутствовать в онтологиях всех актуальных точек зрения.

2. Концепты из остальных компонентов постановки должны быть понимаемы во всех онтологиях, где они не представлены непосредственно.

3. Контекст и поле верификации должны удовлетворять условиям:

- неизбыточности (из сходных концептов оставляется только наиболее информативный);

- непротиворечивости (концепты, объявленные в рамках каких-либо онтологий как взаимоисключающие в составе моделей тех проблем, которые аналогичны данной, удаляются из состава компонента);

- полноты относительно системы точек зрения (все концепты, представленные в данном компоненте постановки в разных онтологиях, кроме нарушающих два первых условия, включаются).

4. Для ЦХ, являющейся концептом, соответствующим результату решения проблемы, в ее модели отображается такое из представлений, имеющихся в онтологиях точек зрения, которое является компромиссным выбором [6].

Стратегия S_2 реализует выявление тех точек зрения, которые нецелесообразно представлять в экспертной группе из-за резко отличающегося концептуального представления ЦХ, и соответствующее формирование экспертной группы.

Стратегия S_3 заменяет процедуру независимого индивидуального оценивания использованием метода комиссий [1]. При этом экспертам заблаговременно предоставляются начальный вариант постановки проблемы и перечень гипотетически конфликтных аспектов в ее составе. Тем самым готовится почва для конструктивных дебатов и формируются аспекты будущего обоснования выработанного коллективного решения.

Стратегия S_4 начинается проведением первого тура индивидуального оценивания при некоторой заданной постановке. Полученные индивидуальные оценки анализируются в совокупности с замечаниями экспертов относительно элементов используемой постановки. Выявляются структурные различия спорных элементов постановки, имеющие место в онтологиях разных точек зрения. Реализуется второй тур экспертизы, в котором носителю точки зрения предоставляется концептуальная версия спорных элементов, которая принадлежит точке зрения, породившей оценку, наиболее отличную от его собственной (в случае понимаемости такой версии в рамках его онтологии). Вместе с тем предоставляется и соответствующая оценка. Такая стратегия является концептуально-ориентированной модификацией метода Дельфи [1,7], позволяющей повысить шансы на сходимость многотурового процесса.

Компьютерная поддержка описанных стратегий аналитического сопровождения экспертизы должна включать решение задач выявления и оценки ряда отношений между концептами. Одним из важнейших среди них является отношение сходства.

В рамках стратегии S_1 анализ такого отношения служит:

- отбору проблем-аналогов из онтологий актуальных точек зрения;
- поиску избыточных элементов в объединенном контексте;
- оценке степени близости концептуальных трактовок ЦХ экспертируемых объектов в онтологиях различных точек зрения (в процедурах поиска концептуального компромисса).

В рамках стратегии S_2 отношение сходства лежит в основе оценки совместимости заданных точек зрения при оценивании заданного класса объектов экспертизы.

Стратегии S_3 и S_4 делают актуальной задачу отбора тех трактовок элементов постановки проблемы, которые максимально отличаются в разных точках зрения (для ознакомления с ними либо членов комиссии, в случае S_3, либо участников второго тура экспертизы, в случае S_4).

Остановимся на особенностях таких онтологий знаний о ПрО экспертизы, которые соответствуют деятельности различных профессиональных и ведомственных групп специалистов и могут успешно использоваться при аналитическом сопровождении процессов экспертного оценивания.

Концептуальная модель (КМ) ПрО, служащая основой такой онтологии, должна обеспечивать следующие возможности:

- представление и использование неполного знания о концептах ПрО;
- сочетание знаний о классификационных и ситуационных соотношениях концептов;
- увязку между собой ракурсов ПрО, связанных с объектами деятельности, решаемыми проблемами, выполняемыми действиями и осуществляемыми коммуникациями.

Специфика класса КМ и рассмотренных выше задач, решение которых должно быть поддержано, определяет синтаксические и семантические требования к разрабатываемому формализму отношения сходства, как это имело место в работах по формализации таких отношений для концептуальных моделей других проблемных ориентаций [8,9].

В качестве модели, построенной с учетом перечисленных выше возможностей описания онтологий, рассмотрим предложенную в [10] КМ знаний о ПрО. В этой КМ способом полного определения концепта является его описание через связи с другими концептами и параметрами, устанавливаемые посредством типизированных n-арных отношений определения.

Пусть K, $D(\kappa)$ и $C(\kappa)$, $\kappa \in K$ – множества соответственно категорий концептов КМ, типов определений для концептов категории κ и концептов категории κ в составе КМ, идентифицированных своими именами;

$CC = \cup_{\kappa \in K} C(\kappa)$ и $DD = \cup_{\kappa \in K} D(\kappa)$ – множества всех концептов КМ и типов их определений;

T^0 – выделенный тип определения, реализующего связь концептов КМ с независимыми параметрами PAR, множество которых определено для данной ПрО;

$T \in D(\kappa)$ – некоторый тип определений концептов категории κ;

$x \in C(\kappa)$ и $T(x)$ – имя некоторого концепта категории κ и его определение типа T;

cat: $CC \rightarrow K$ – функция, сопоставляющая концепту его категорию.

С формальной точки зрения полное определение $Def(c)$ в КМ концепта $c\in C(\kappa)$, $\kappa\in K$, есть конъюнкция его определений $T(c)$, обладающих типами $T\in D(\kappa)$, количество членов которой совпадает с количеством типов $|D(\kappa)|$. При этом $T(c)$ и $T^0(c)$ представляют собой пары

$$T(c) = \langle L(c),S(c), \rangle; T^0(c) = \langle L(c), S^0(c), \rangle, \quad (1)$$

где $L(c)\in\{d,p,u,n\}$ – параметр, посредством которого определения $T(c)$ и $T^0(c)$ могут идентифицироваться как полностью определенные (d), неполностью определенное (p), неизвестное (u) и неактуальное для данного концепта (n);

$$\begin{aligned} &S(c) = \langle B(c,S); A(c,S); I(c,S) \rangle, S^0(c) = \\ &=\langle B(c,S^0); I(c,S^0)\rangle, \text{ если } L(c)\in\{d,p\}, \quad (2)\\ &S(c), S^0(c) = \theta, \text{ если } L(c)\in\{u\}; \end{aligned}$$

θ - специальный символ, обозначающий отсутствие знаний о некотором определении.

Использование параметра $L(c)$ позволяет различать знания, закрытые для развития ($L(c)\in\{d,n\}$), и, соответственно, открытые ($L(c)\in\{p,u\}$), а также отличать случай незнания связей, соответствующих данному типу определения ($L(c)=u$), от случая знания о неактуальности определения данного типа для данного концепта ($L(c)=n$).

В выражении (2)

$$\begin{aligned} &B(c,S) = \{c_i', i=1,\dots,N(c,S)\}; D(cat(c_i'))= \\ &=D(cat(c_j')) , i,j \geq 1; \\ &A(c,S) = \{\langle S'(c_i'), ASB(c,c_i',S')\rangle, S' \in \\ &\in D(cat(c_i'))\rangle, c_i'\in B(c,S)\}, \quad (3)\\ &ASB(c,c_i',S') = \\ &= \{c_j''\in B(c_i',S'), j=1,\dots,N(c',S')\}; \\ &B(c,S^0)=\{p_i\in PAR, i\leq |PAR|\}. \end{aligned}$$

Таким способом определение $T(c)$ концепта c задается (посредством *базиса* $B(c,S)$) через другие концепты, являясь n-арным отношением над ними с инвариантами $I(c,S)$. При этом *актуальное раскрытие* $A(c,S)$ позволяет задать для любого концепта c', входящего в базис, ту совокупность определений концепта c', которая принимается во внимание, и то подмножество концептов из базисов этих (косвенно используемых для $Def(c)$) определений, которое учитывается. Будем говорить о таких определениях и концептах, что они *актуальны* для концепта c. Формулировки "учитывается" и "принимается во внимание" для актуальных концептов формально означают их используемость в качестве параметров в функциях и процедурах, составляющих множество $I(c,S)$.

Функции из множества $I(c,S)$ являются предикатами, которые определяют необходимые требования к экземплярам концептов $c'\in B$, включаемым в экземпляр c посредством метаотношения S. В то же время процедуры $PROC : C_{in} \Rightarrow R_{out}$, включаемые в состав некоторых типов определений, задают способ конструирования набора значений для параметров либо для экземпляров концептов R_{out}, принадлежащих базису другого определения c, на основании множества экземпляров концептов C_{in}, входящих в базис текущего определения.

Каждому типу определения S сопоставим аксиому симметрии

$$\begin{aligned} &\forall c,d\in CC, S\in D(cat(c)) (d \in B(c,S)) \Rightarrow \\ &\Rightarrow (\exists S^*\in D(cat(d))\, c\in B(d,S^*)). \quad (4) \end{aligned}$$

Определения S и S^* будем называть *сопряженными*.

Формализация отношения сходства. Прежде чем перейти к формальному конструированию, исследованию и метризации отношения сходства между концептами описанной КМ, отметим некоторые важные содержательные особенности.

Использование аппарата актуального раскрытия элементов базиса в определениях концептов открывает путь для введения трех качественно различных парадигм тождества концептов c_1 и c_2, определяемых по заданному типу их определения S. Во всех случаях будем говорить о структурном тождестве определений (1)-(3), игнорируя их инварианты. Такой же подход будет сохранен и при рассмотрении отношений сходства. Эта ограниченность оправдана первоочередной целью аналитического сопровождения экспертизы: добиться, чтобы носители различных точек зрения, оперируя некоторым понятием из постановки проблемы, рассматривали его с позиции взаимосвязей с элементами одной и той же совокупности понятий, проблем, документов и коммуникаций в ПрО экспертизы. Это позволяет улучшить шансы сопоставимости оценок, которые будут получены. Распространение же требований на свойства этих взаимосвязей привело бы к выхолащиванию самой сути экспертного метода.

В рамках первой парадигмы тождественными полагаются концепты с совпадающими базисами определений, т.е. определенные, по типу S, через одни и те же концепты. Применительно к паре КМ, отображающих различные точки зрения на ПрО, такое тождество игнорирует различия в понимании концептов базиса, имеющиеся между этими точками зрения. Эту парадигму можно назвать *тождеством по базису(В-парадигмой)*.

Вторая парадигма требует для тождественности понимания концепта C при разных точках зрения на ПрО наличия, помимо тождества между базисами, еще и тождества их актуальных раскрытий. Назовем ее *тождеством по комплексному основанию* (*W –парадигмой*).

Наконец, в рамках третьей парадигмы, для тождества двух пониманий C (относительно определения типа S) безразлично, одинаковые ли концепты задекларированы в базисе. Важно, чтобы их актуальные раскрытия были одинаковыми. Данную парадигму назовем тождеством *по актуальному раскрытию* (*А-парадигмой*).

Для определения $T(a)$ концепта $a\in C(\kappa)$ выделим его *семантическое поле* $SF(a,S)$. Определим его как связный ориентированный ациклический граф (дерево) с корнем a

$$SF(a,S) = \langle LL,BB \rangle, LL = \cup_{k=1,\dots,M}L_k, BB = \\ = \cup_{k=1,\dots,M}B_k, L_k = \{v_{ki}, i\geq 1\}, B_k = \{b_{ki}, i\geq 1\},$$

где LL – множество вершин; BB – множество дуг; M – количество уровней дерева;
L_k, v_{ki}, B_k – соответственно k-й уровень; i-я вершина этого уровня; множество дуг b_{ki} между вершинами $(k\text{-}1)$-го уровня и подчиненными им вершинами k-го уровня.

Вершину v_{ki} k-го уровня $SF(a)$, подчиненную вершине $(k\text{-}1)$-го уровня $v_{(k-1)j}$, определим как кортеж, четыре элемента которого представляют собой имена:

- типа некоторого определения, которым обладает концепт $y^{(k-1)}$, стоящий на четвертом месте в кортеже для вершины $v_{(k-1)j}$, подчиняющей v_{ki} (S^k);
- такого концепта x^k из базиса определения S^k концепта $y^{(k-1)}$, что по крайней мере некоторые его определения актуальны для $y^{(k-1)}$;
- типа некоторого определения x^k, актуального для $y^{(k-1)}$ ($S^k{}_u$);
- концепта y^k из базиса определения $S^k{}_u$ концепта x^k, актуального для $y^{(k-1)}$.

В качестве листьев $SF(a,S)$ используем такие кортежи v_{Mi}, в которых четвертый элемент y^M либо представляет собой параметр, либо же содержит концепт y^k, $k<M$, опосредованным раскрытием которого по актуальности с позиций соответствующего определения SS он является в базисе своего актуального определения SS^*, причем SS^* и SS сопряжены по аксиомам КМ ПрО в смысле (4).

Сохраняя нотацию (2)-(3), формально опишем вид вершин графа $SF(a,S)$:

$$\begin{aligned}&v_{1i} = \langle S^1{}_l, x^1{}_u, S^1{}_{uv}, y^1{}_{uvl}\rangle;\ v_{ki} = \\ &= \langle S^k{}_q, x^k{}_p, S^k{}_{pr}, y^k{}_{prw}\rangle;\ v_{Mi} = \\ &= \langle S^M{}_m, x^M{}_n, S^M{}_{nt}, y^M{}_{nts}\rangle, \\ &S^1{}_l = S, l\geq 1, y^0{}_w = a, w=1;\end{aligned} \tag{5}$$

$$\begin{aligned}&S^k{}_q\in D(cat(y^{(k-1)}{}_w)), x^k{}_p\in B(x^{(k-1)}{}_u, S^k{}_q), \\ &S^k{}_{pr}\in D(cat(x^k{}_p)), \\ &y^k{}_{prw}\in ASB(x^k{}_p, S^k{}_{pr}), k=1,\dots,M; \\ &\forall v_{Mi}\ ((S^M{}_m = S^0) \wedge (x^M{}_n \in PAR) \wedge \\ &\wedge (S^M{}_{nt} = y^M{}_{nts} = \varpi)\ ! \\ &!(\exists y^k{}_{prw}, SS\in D(cat(y^k{}_{prw}))\ | \\ &|(S^M{}_m = SS^*)\wedge(\langle SS, x^{(k+1)'}, \\ &S^{(k+1)'}, y^{(k+1)'}\rangle \in L_{(k+1)})\wedge \\ &\wedge(y^k{}_{prw}\in B(S^M{}_m, y^M{}_{nts}),\end{aligned} \tag{6}$$

где ϖ - специальный символ, обозначающий неактуальность некоторых элементов нотации для листьев семантического поля.

Будем писать: $(x,m)\in KOR$, если x – m-й элемент некоторого кортежа KOR. Предположим, что на множестве CC определено бинарное отношение SKF^T синтаксического конфликта с позиций произвольного определения $T=(L,S)\in DD$. В дальнейшем рассмотрим такое подмножество $C^T(\kappa)\subseteq C(\kappa)$, что

$$\forall a,b\in C^T(\kappa),\ \forall(x,y)|\ (x,u)\in v_{ki}\in SF(a,S),$$
$$(y,u)\in w_{pq}\in SF(b,S)\ \neg SKF^T(x,y),u\in\{2,4\}).$$

Опираясь на введенную выше нотацию частичного определения (2)-(3) и описанные парадигмы тождества концептов, введем три парадигмы π сходства с позиций определения T: по базису (π=B), по актуальному раскрытию (π=A) и по комплексному основанию (π=W). Каждой парадигме сопоставим два типа сходства: прямое и косвенное, а также две градации силы сходства: сильное и слабое сходство.

Определение 1. Два концепта полагаются связанными отношением *прямого сильного сходства* по T в парадигме π (DSS^T_π, $\pi\in\{B,W,A\}$):

- *по базису* (π=B), если базисы сопоставляемых концептов имеют общие элементы;

- *по комплексному основанию* (π=W), если пересечения их базисов и подмножеств актуальных раскрытий, соответствующих элементам пересечения, одновременно не пусты;

- *по актуальному раскрытию* (π=A), если у элементов базиса определения T сопоставляемых концептов имеются общие типы определений и, кроме того, актуальные раскрытия базисов таких элементов имеют непустое пересечение.

Формальные выражения для трех введенных форм отношения сходства имеют вид

$$\begin{aligned}&\forall a,b\in C^T(\kappa)\quad DSS^T_B(a,b)\Leftrightarrow\\&\Leftrightarrow B(a,S)\cap B(b,S)\neq\varnothing,\\&DSS^T_W(a,b)\Leftrightarrow\exists(x\in B(a,S)\cap B(b,S),\\&V\in D(cat(x))\ |\ ASB(a,x,V)\cap\\&\cap ASB(b,x,V))\neq\varnothing),\\&DSS^T_A(a,b)\Leftrightarrow\exists(x\in B(a,S),\ y\in B(b,S),\\&V\in D(cat(x))\cap D(cat(y))\\&|ASB(a,x,V)\cap ASB(b,y,V)\neq\varnothing.\end{aligned}\tag{7}$$

Определение 2. Два концепта полагаются связанными отношением *прямого слабого сходства* по T в парадигме π (DWS^T_π, $\pi\in\{B,W,A\}$):

- *по базису* (π=B), если существует хотя бы одна пара элементов базисов сопоставляемых концептов, для которой имеет место DSS^R_B;

- *по комплексному основанию* (π=W), если существует хотя бы одна пара элементов базисов сопоставляемых концептов и одновременно элементов из актуальних раскрытий базисов этих последних, для которой имеет место DSS^R_W;

- *по актуальности* (π=A), если в актуальных раскрытиях базисов сопоставляемых концептов имеет место отношение DSS^R_A.

Формальное определение трех отношений DWS^T_π задается следующими выражениями:

$$\begin{aligned}&\forall a,b\in C^T(\kappa)\ DWS^T_B(a,b)\Leftrightarrow\\&\Leftrightarrow(\exists(x\in B(a,S),y\in B(b,S),\\&R\in D(cat(x))\cap D(cat(y))=E,\\&ASB(a,x,R)\in A(a,S),\\&ASB(b,y,R)\in A(b,S)|\ DSS^R_B(x,y);\\&DWS^T_W(a,b)\Leftrightarrow\\&\Leftrightarrow(\exists(x\in B(a,S),y\in B(b,S),R\in E;\\&ASB(a,x,R)\in A(a,S);\\&ASB(b,y,R)\in A(b,S))|\\&|(DSS^R_W(x,y))\wedge\\&\wedge(\exists(u\in ASB(a,x,R);\\&v\in ASB(b,y,R)|\ DSS^R_W(u,v));\\&DWS^T_A(a,b)\Leftrightarrow(\exists(x\in B(a,S),y\in B(b,S);\\&R\in E,ASB(a,x,R)\in A(a,S);\\&ASB(b,y,R)\in A(b,S))|(\exists(u\in ASB(a,x,R);\\&v\in ASB(b,y,R))|DSS^R_A(u,v)).\end{aligned}\tag{8}$$

Определение 3. Два концепта полагаются связанными отношением *косвенного сильного* (соответственно, *слабого) сходства* по T в парадигме π (соответственно ISS^T_π либо IWS^T_π), $\pi\in\{B,A,W\}$, если семантические поля этих концептов содержат на одном и том же уровне k, кроме первого, такие вершины v' и v'' вида (5), четвертые элементы в описаниях которых связаны отношением прямого сильного (соответственно, слабого) сходства в парадигме π по всем определениям, одновременно актуальным для концептов $y^{(k-1)\prime}_w$ и $y^{(k-1)\prime\prime}_r$, представляющих собой четвертые элементы вершин предыдущего уровня, которым вершины v' и v'' непосредственно подчинены в соответствующих семантических полях.

Уровень k назовем *реализующим сходство* уровнем.

С формальных позиций категории косвенного сходства определяются следующими выражениями (с использованием тех же обозначений, что и в выражениях (5)-(6)):

$$\begin{aligned}
&\forall a,b\in C^T(\kappa)\quad ISS^T_\pi(a,b) \Leftrightarrow\\
&\Leftrightarrow (\exists U\in D(cat(y^{(k-1)\prime}_w)) \cap\\
&\cap D(cat(y^{(k-1)\prime\prime}_r))) \wedge (\exists(k>1;\\
&u_{ki}=\langle S^k_q, x^k_p, S^k_{pr}, y^k_{prw}\rangle\in SF(a,S);\\
&v_{kj}=\langle S^k_t, x^k_l, S^k_{lm}, y^k_{lmn}\rangle\in SF(b,S))| \qquad (9)\\
&|DSS^U_\pi(y^k_{prw}, y^k_{lmn}));\\
&IWS_\pi(a,b)\Leftrightarrow(\exists U\in G)\wedge\\
&\wedge (\exists(k>1,u_{ki}=\langle S^k_q,x^k_p,S^k_{pr},y^k_{prw}\rangle\in SF(a,S);\\
&v_{kj}=\langle S^k_t,x^k_l,S^k_{lm},y^k_{lmn}\rangle\in SF(b,S))|\\
&|DWS^U_\pi(y^k_{prw}, y^k_{lmn})).
\end{aligned}$$

Свойства отношения сходства. Непосредственно из определений вытекают следующие свойства введенных отношений сходства.

Лемма 1. Все отношения сходства (7)-(9) представляют собой толерантности.

Замечание 1. В предельном случае, когда в выражениях (7) вместо пересечения используемых в нотации множеств имеет место их тождество, отношения прямого сильного сходства приобретают свойство транзитивности, превращаясь таким образом в эквивалентности.

Замечание 2. В предельном случае, когда выражения (8) заменены более сильными условиями:

$$\begin{aligned}
&\forall a,b\in C^T(\kappa)\quad DWS^{*T}_B(a,b)\Leftrightarrow\\
&\Leftrightarrow (\forall(x\in B(a,S))\ \exists(y\in B(b,S)))\wedge\\
&\wedge(\forall(y\in B(b,S))\ \exists(x\in B(a,S)))|\\
&|\ \exists R\in D(cat(x))\cap D(cat(y))=E;\\
&ASB(a,x,R)\in A(a,S);\\
&ASB(b,y,R)\in A(b,S)|DSS^R_B(x,y)); \qquad (10)\\
&DWS^{*T}_W(a,b)\Leftrightarrow(\forall(x\in B(a,S))\\
&\ \exists(y\in B(b,S)))\wedge(\forall(y\in B(b,S))\\
&\ \exists(x\in B(a,S)))|\exists R\in D(cat(x))\cap\\
&\cap D(cat(y))=E,\ ASB(a,x,R)\in A(a,S),\\
&ASB(b,y,R)\in A(b,S)|(DSS^R_W(x,y))\wedge\\
&\wedge(\forall(u\in ASB(a,x,R))\exists(v\in ASB(b,y,R))\wedge\\
&\wedge(\forall(v\in ASB(b,y,R))\exists(u\in ASB(a,x,R))|\\
&|DSS^R_W(u,v));\\
&DWS^{*T}_A(a,b)\Leftrightarrow(\forall(x\in B(a,S))\\
&\exists(y\in B(b,S)))\wedge(\forall(y\in B(b,S))\ \exists(x\in B(a,S)))|\\
&|\ \exists R\in D(cat(x))\cap D(cat(y))=E,\\
&ASB(a,x,R)\in A(a,S),ASB(b,y,R)\in A(b,S)|\\
&\wedge(\forall(u\in ASB(a,x,R))\exists(v\in ASB(b,y,R))\wedge\\
&\wedge(\forall(v\in ASB(b,y,R))\exists(u\in ASB(a,x,R))|\\
&|DSS^R_A(u,v)),
\end{aligned}$$

а используемое в (8) отношение DSS^R_π, $\pi\in\{B,W,A\}$ представляет собой эквивалентность, отношения прямого слабого сходства также приобретают свойство транзитивности, превращаясь таким образом в эквивалентности.

Соотношения между формами введенных отношений сходства устанавливает

Лемма 2. Справедливы следующие утверждения:

а) сильное сходство по комплексному основанию влечет одновременное наличие сходства по актуальности и по базису. Обратное справедливо только в тех случаях, когда в выражениях (7) для DSS^T_A $x=y$ либо в выражении (8) для DWS^T_A дополнительно имеет место $DSS^R_W(x,y)$, где R – общее актуальное определение сопоставляемых концептов, либо же указанные соотношения справедливы для x и y, реализующих косвенное сходство:

$$\begin{aligned}
&\forall a,b\in C^T(\kappa)\ DSS^T_W(a,b)\Rightarrow DSS^T_A(a,b)\wedge\\
&\wedge DSS^T_A(a,b);\ DWS^T_W(a,b)\Rightarrow DWS^T_A(a,b)\wedge\\
&\wedge DWS^T_A(a,b);\\
&\quad ISS^T_W(a,\wedge b)\Rightarrow ISS^T_A(a,b)\wedge ISS^T_B(a,b);\\
&\quad IWS^T_W(a,b)\Rightarrow IWS^T_A(a,b)\wedge IWS^T_B(a,b).
\end{aligned}$$

б) сходство по актуальности не влечет сходства по базису и наоборот. Сходство по базису вытекает из сходства по актуальности в случаях, описанных в утверждении а).

Соотношения между типами введенных отношений сходства устанавливает

Лемма 3. Прямое сильное сходство в парадигме $\pi,\pi\in\{B,A,W\}$ влечет прямое слабое сходство в той же парадигме. Это утверждение справедливо и для косвенного сходства:

$$\forall a,b \in C^T(\kappa) DSS^T{}_\pi(a,b) \Rightarrow DWS^T{}_\pi(a,b);$$
$$ISS^T{}_\pi(a,b) \Rightarrow IWS^T{}_\pi(a,b).$$

Оценка степени сходства. Для решения поставленных выше задач необходима метризация введенных отношений сходства. Перспективным путем ее осуществления представляется использование для оценки степени сходства концептов функции $\mu(SIM)$ с параметром $SIM \in \{DSS^T{}_\pi, DWS^T{}_\pi, ISS^T{}_\pi, IWS^T{}_\pi, \pi \in \{W,A,B\}\}$, удовлетворяющей следующим требованиям.

Т1. $\mu(SIM):SM(SIM,\kappa) \to (0;1]$, $SM(SIM,\kappa)=\{(x',x'')|(x',x'' \in C^T(\kappa)) \wedge \wedge SIM(x',x'')\}$.

Т2. $1- \mu(SIM)$ является метрикой на $SM(SIM,\kappa)$.

Для метризации отношений прямого сильного сходства воспользуемся метрикой [11] на некотором множестве деревьев TREE, использующей только множества их листьев:

$$d(T_1,T_2) = |FV_1 \setminus FV_2|/|FV_1| + |FV_2 \setminus FV_1|/|FV_2| + |FV_2 \cap FV_1| \, ||FV_1|^{-1}-|FV_2|^{-1}|, \qquad (11)$$

где FV_i – множество листьев дерева $T_i \in$TREE, $i=1,2$;

$|A|$ – обозначение мощности множества A.

Семантическому полю $SF(x,S)$, $x \in C^T(\kappa)$ сопоставим два вспомогательных конструкта.

Определение 4. *А*-конструктом уровней $k_1,\ldots,k_p$, $p \geq 1$ для $SF(x,S)$ назовем дерево $ASF(x,S; k_1,\ldots,k_p)$, в котором каждое из множеств $\{v_{ui} = \langle S^u{}_q, x^u{}_p, S^u{}_{pr}, y^u{}_{prw}\rangle, i \geq 1\}$ вершин *u*-го уровня $SF(x,S)$ вида (5),(6), подчиненных вершине $v_{(u-1)j}$, $u=k_1,\ldots,k_p$, заменено вершиной вида $\alpha^*{}_{ue} = \langle S^u{}_q, S^u{}_{pr}, y^u{}_{prw}\rangle$, по-прежнему подчиненной $v_{(u-1)j}$.

Определение **5**. *В*-конструктом уровней $k_1,\ldots,k_p$, $p \geq 1$ для $SF(x,S)$ назовем дерево $BSF(x,S; k_1,\ldots,k_p)$, в котором каждое из множеств $\{v_{ki} = \langle S^k{}_q, x^k{}_p, S^k{}_{pr}, y^k{}_{prw}\rangle, i \geq 1\}$ вершин u-го уровня $SF(x,S)$ вида (5),(6), подчиненных вершине $v_{(u-1)j}$, $u=k_1,\ldots,k_p$, заменено вершиной вида $\beta^*{}_{ue} = \langle S^u{}_q, S^u{}_{pr}\rangle$, по-прежнему подчиненной $v_{(u-1)j}$.

Теорема 1. Функция для оценки степени прямого сильного сходства, удовлетворяющая требованиям Т1,Т2, имеет вид

$$\forall (a,b) \in SM(DSS^T{}_W, \kappa) \quad \mu(DSS^T{}_W; a,b) = 1-0.5\{|L_1(SF(a,S)) \setminus L_1(SF(b,S))|/|L_1(SF(a,S))| + |L_1(SF(b,S)) \setminus L_1(SF(a,S))|/|L_1(SF(b,S)|| + |L_1(SF(a,S)) \cap L_2(SF(b,S)| \times ||L_1(SF(a,S))|^{-1}-|L_1(SF(b,S))|^{-1}|\};$$

$$\mu(DSS^T{}_A; a,b) = 1-0.5\{|L_1(ASF(a,S;1)) \setminus L_1(ASF(b,S;1))|/|L_1(ASF(a,S;1))| + |L_1(ASF(b,S;1)) \setminus L_1(ASF(a,S))|/|L_1(ASF(b,S;1)|| + |L_1(ASF(a,S;1)) \cap L_2(ASF(b,S;1)| \, ||L_1(ASF(a,S;1))|^{-1}-|L_1(ASF(b,S;1))|^{-1}|\}; \qquad (12)$$

$$\mu(DSS^T{}_B;a,b) = 1-0.5\{|B(a,S) \setminus B(b,S)|/|B(a,S)| + |B(b,S) \setminus B(a,S)|/|B(b,S| + |B(a,S) \cap B(b,S)| \, ||B(a,S)|^{-1}-|B(b,S)|^{-1}|\}.$$

Выражения (12) непосредственно вытекают из требований Т1, Т2 и выражения (11) при выборе в качестве T_i, $i=1,2$, одноуровневых деревьев, листьями которых являются частично совпадающие (по определению прямого сильного сходства) элементы соответственно семантических полей сопоставляемых концептов и их *А*-конструктов первого уровня и базисов.

Остальные формы сходства метризуем с помощью расстояния Хемминга ρ^h, аксиоматически введенного на множестве TREE К.Богартом [11] и учитывающего, прежде всего, различия в структуре деревьев, соответствующих сопоставляемым концептам. Обозначим:

$UN=\{\varepsilon_v, v=1,\ldots,H\}$ – множество всех вершин деревьев из множества TREE;

$MS_i=|v_{iuv}|_{u,v=1,\ldots,H}$ – матрица, задающая дерево $T_i \in$TREE, с элементами: $v_{iuv} = 1$, если ε_v подчинена ε_u в дереве T_i; $v_{iuv} = -1$, если ε_u подчинена ε_v в дереве T_i; $v_{iuv} = 0$, если ε_v не подчинена ε_u и ε_u не подчинена ε_v в дереве T_i.

По определению [11]

$$\rho^h(T_i, T_j) = \Sigma_{1 \leq u < v \leq H} |v_{iuv} - v_{juv}|. \qquad (13)$$

Теорема 2. Функция для оценки степени сходства в форме $SIM \in \{DWS^T{}_\pi, ISS^T{}_\pi, IWS^T{}_\pi, \pi \in \{W,A,B\}\}$, удовлетворяющая требованиям Т1,Т2, имеет вид

$$\forall(a,b)\in SM(SIM,\kappa) \quad \mu(SIM; a,b) = = 1 - 0.5\rho^h(T^{SIM}(a),T^{SIM}(b)), \qquad (14)$$

где $\forall x\in\{a,b\}$ $T^{SIM}(x)=BSF(x,S;2)$, *если* $SIM=DWS^T{}_B$;
$T^{SIM}(x)=SF(x,S)$, *если* $SIM\in\{DWS^T{}_W, ISS^T{}_W, IWS^T{}_W\}$;
$T^{SIM}(x)=ASF(x,S;2)$, *если* $SIM=DWS^T{}_A$;
$T^{SIM}(x) = BSF(x,S;k_1,\ldots, k_p)$, если $SIM = ISS^T{}_B$ и сходство реализуют уровни $k_1,\ldots, k_p$;
$T^{SIM}(x)=BSF(x,S;(l_1+1),\ldots,(l_q+1))$, если $SIM= =IWS^T{}_B$ и сходство реализуют уровни $l_1,.., l_q$;
$T^{SIM}(x)= ASF(x,S;m_1,\ldots, m_r)$, если $SIM=ISS^T{}_A$ и сходство реализуют уровни $m_1,\ldots, m_p$;
$T^{SIM}(x)=ASF(x,S;(n_1+1),\ldots,(n_s+1))$, если $SIM=IWS^T{}_B$ и сходство реализуют уровни $n_1,\ldots, n_s$.

Выражение (14) вытекает из выражения (13) при подстановке в него, вместо T_i, T_j, семантических полей сопоставляемых концептов либо соответствующих им конструктов, вершины которых частично совпадают по определению сходства формы *SIM*.

Замечание 3. При сформулированных в замечании 1 и 2 условиях, когда отношения прямого сходства становятся эквивалентностями, степень их сходства равна 1.

Отметим естественное свойство введенных функций степени сходства (12),(14), непосредственно вытекающее из выражений (12),(14) и определений форм сходства.

Лемма 4. При пополнении базиса частично определенного концепта элементами базиса полностью определенного концепта степень их сходства не убывает, а при пополнении его элементами, не входящими в этот базис, – соответственно не возрастает:

$$\forall(SIM\in\{DSS^T{}_\pi, DWS^T{}_\pi, ISS^T{}_\pi, IWS^T{}_\pi, \pi\in\{W,A,B\}\}, (a,b)\in SM(SIM,\kappa)|\ L(a) = = d,L(b) =p) \qquad (15)$$
$$\mu(SIM; a,b_*) \le \mu(SIM; a,b) \le \le\mu(SIM; a,b^*),$$

где $L(b_*)=L(b^*)=d$; $B(b_*,S)=B(b,S)\cup\Delta_*$, $\Delta_*\subseteq B(a,S)$; $B(b^*,S)=B(b,S)\cup\Delta^*$, $\Delta^*\not\subset B(a,S)$.

Сформулированная лемма позволяет сопоставить паре концептов с различной степенью определенности специальную *характеристику* их сходства в виде тройки

$$CHAR(SIM; a,b) = \langle \mu(SIM; a,(b\backslash a)); \mu(SIM; a,b); \mu(SIM; a,(b\cup a))\rangle, \qquad (16)$$

где (b\a) – концепт, базис определения типа *T* которого получен в результате пополнения соответствующего базиса концепта *b* всеми концептами, допустимыми согласно аксиомам КМ ПрО, не принадлежащими базису *B(a,S)* концепта *a* и не связанными отношением синтаксического конфликта SKF^T с элементами базиса *B(b,S)* концепта *b*;
$(b\cup a)$ – концепт, базис определения *T* которого получен вседедствие пополнения базиса концепта *b* всеми элементами *B(a,S)*, не принадлежащими *B(b,S)*.

Использование отношения сходства при аналитическом сопровождении экспертиз. Выше формализованы, метризованы и исследованы двенадцать форм отношения сходства между концептами онтологий экспертных точек зрения, описанных посредством КМ специального вида. Они ориентированы на поддержку сопоставления и отбора концептов для формирования постановки проблемы экспертного оценивания в условиях множественности концептуально различных точек зрения на ПрО экспертизы.

Четыре описанные стратегии аналитического сопровождения экспертизы определяют в качестве сфер использования отношения сходства решение задач поиска на множестве концептов элемента либо группы элементов с такими свойствами:

- наименее сходных с остальными;
- имеющих наименьшее (в среднем) отличие от остальных элементов;
- наиболее сходных с выделенным элементом множества.

Перечисленные задачи могут решаться при двух различных целевых установках:

- выявить все элементы, обладающие одним из указанных свойств;
- найти среди выявленных элементов тот, для которого свойство наиболее выражено.

В связи с наличием в используемой

КМ аппарата представления неполных знаний о концептах характеристика сходства формы SIM между концептами *a*,*b*, имеющими различную степень определенности знаний, была представлена тройкой *CHAR(SIM; a,b)* (16). Ее центральный элемент задает оценку степени сходства при текущем состоянии знаний, а первый и третий соответствуют предельной степени сходства при развитии знания, соответственно, в неблагоприятном и в благоприятном для сходства концептов направлении.

Благодаря этому каждая из перечисленных задач при наличии в анализируемом множестве концептов с разными степенями определенности знаний может иметь тройку решений, элементы которой соответствуют принятию одного из трех предположений одновременно для всех частично определенных концептов множества.

Поскольку для всех форм отношения сходства дополнение до единицы функции оценки степени сходства удовлетворяет аксиомам метрики, то любая из этих форм может использоваться при решении перечисленных задач.

Выбор форм отношения сходства для решения конкретной задачи является выбором нужных значений для таких параметров, как парадигма сходства (*B*, *A* или *W*), тип сходства (прямое или косвенное) и сила сходства (сильное или слабое).

Выбор парадигмы сходства определяется в первую очередь предположениями о системе онтологий, использование которой планируется при решении задачи.

B-сходство предпочтительно для использования, если известно (по договоренности или на основе опыта предыдущих экспертиз), что концепты из состава базиса принадлежат прагматически выделенному множеству концептов, одинаково понимаемых в онтологиях всех учитываемых точек зрения.

A-сходство может использоваться, когда между онтологиями точек зрения ожидается наличие соответствий, порождающих пары разных концептов, тождественных (или слабо различающихся) в той части их определений, которая актуальна для сравниваемых концептов. Имеется в виду, что при сравнении концептов c_1 и c_2, в базисы определений которых входят соответственно x_1 и x_2, для x_1 и x_2 актуальны те части их определений, которые одинаковы или сходны.

Дополнительным при этом является предположение такого же характера, как и сформулированное выше для *B*-сходства, но отнесенное не к элементам базиса, а к элементам их актуального раскрытия. Это предположение не требуется, если известно, что инварианты отношений определения не используют более глубоких концептуальных раскрытий концептов базиса (как, например, в КМ KAOS [12]).

W-сходство предпочтительно для использования в случае, если знания о сравниваемых концептах неполны и в ходе их развития допускается возможность расширения множества актуальных аспектов определения концептов базиса.

Выбор между прямым и косвенным сходствами зависит от прагматики решаемой задачи. Для случая, когда анализ сходства используется как промежуточный шаг при оценке понимаемости концептов другой онтологии, предпочтителен тип прямого сходства. В случае использования этого отношения для анализа информативности одного концепта по отношению к другому предпочтительно косвенное сходство.

Такой параметр, как сила сходства, варьируется при построении процедур сравнения для решения перечисленных выше задач поиска. Так, при поиске концепта, наиболее сходного с выделенным, можно использовать следующую последовательность действий.

1. Отбор элементов по наличию прямого сильного сходства (в множество Z_1).

2. Если $Z_1=\varnothing$, то отбор по наличию косвенного сильного сходства (в множество Z_2). Иначе $Z= Z_1$ и переход к действию 5.

3. Если $Z_2=\varnothing$, то отбор по наличию прямого слабого сходства (в множество Z_3). Иначе $Z= Z_2$ и переход к действию 5.

4. Если $Z_3=\varnothing$, то решение отсутствует и процесс его поиска окончен. Иначе $Z= Z_3$.

5. Выбор из множества Z элемента A, имеющего максимальное значение степени сходства соответствующей формы.

Из этого примера видно, как в рамках решения одной задачи могут использоваться разные формы отношения сходства, различающиеся типом и силой. Это предоставляет возможность для реализации предпочтений, которыми руководствуется аналитик.

Выводы

Аналитическое сопровождение экспертизы преследует цель получения обобщаемых и обоснованных решений в условиях привлечения представителей концептуально различных точек зрения на ПрО принятия решений.

Формализация отношения сходства между концептами онтологий экспертных точек зрения является необходимым условием решения задач аналитического сопровождения экспертизы.

Предложенные в статье формы отношения сходства и их метризации ориентированы на работу с неполным знанием и учитывают специфику концептуальных моделей экспертных точек зрения.

Предлагаемый формализм отношения сходства может быть использован в процедурах обеспечения аналитического сопровождения экспертизы в среде систем комплексной поддержки принятия решений.

1. *Крымский С.Б., Жилин Б.Б., Паниотто В.И.* и др. Экспертные оценки в социологических исследованиях. – К.: Наук. думка, 1990. – 320 с.

2. *Наппельбаум Э.Л., Поспелов Д.А.*Проблемы коллективных решений и экспертных оценок // Вопросы кибернетики. Теория принятия решений. – 1975. – Вып.8. – С. 86-102.

3. *Экспертные* оценки. Методы и применение: (Обзор) / Д.С.Шмерлинг, С.А.Дубровский, Т.Д.Аржанова, А.А.Френкель // Статист. методы анализа экспертных оценок: Уч. зап. по статистике. – М.: Наука, 1977. – **29**. – 384 с.

4. *Ильина Е.П.* Экспертная методология в информационно-аналитических системах // Пробл. программирования. – 2001. – № 1-2. – С.13-22.

5. *Программно-целевое* управление оборонным планированием при реформировании вооруженных сил. Методологические основы и перспективы автоматизированной поддержки. / Е.П Ильина., И.П.Синицын, О.А Слабоспицкая., В.Ю. Суслов, Е.Ф. Шелест Т.Л, Яблокова // Київ: Наук. думка, 2004. – 172с.

6. *Ильина Е.П., Слабоспицкая О.А.* Концептуальный компромисс: интеграция экспертных точек зрения // Матеріали 12-ї Міжнар. конф. з автоматичного управління, Харків: НТУ "ХПІ", 30 травня – 3 червня 2005р. – С.100-101.

7. *Литвак Б.Г.* Экспертные оценки и принятие решений. – М.: Патент, 1996. – 275 с.

8. *Spanoudakis G., Constantopoulus P.* Integrating Specifications: A Similarity Reasoning Approach // CIKM 99. – P.212-225.

9. *Palopoli L., Sacca D., Ursino D.* An automatic technique for detecting type conflicts in database schemes // CIKM 98. – P.306-313.

10. *Ильина Е.П.* Методы представления и комплексного использования структур знаний различных уровней формализации в описании экспертной точки зрения на предметную область решаемой проблемы. // Пробл. программирования. – 2002. – №1-2. – С.409-420.

11. *Раппопорт А.М., Шнейдерман М.В.* Анализ экспертных суждений, заданных в виде структур // Прикладной многомерный статистический анализ: Уч. зап. по статистике. Т. 33 – М.: Наука, 1978. – 392 с.

12. *Van Lamsweerde A., Darimont R., Letier E. Managing* Conflicts in Goal-driven Requirements Engineering // IEEE Trans. on Software Eng.: Special Issue on Inconsistency Management in Software Development. – 1998, – **24**, No. 11. – P. 908-926.

Получено 01.07.05

Об авторах

Ильина Елена Павловна

канд. физ.-мат. наук,
вед. науч. сотрудник

Слабоспицкая Ольга Александровна,

науч. сотрудник

Место работы авторов:

Институт программных систем
НАН Украины
03680, Киев-187, пр. Акад. Глушкова, 40
Тел. (044) 526 4579