

$$I = \log_2 L^p = \log_2 \sum_{m=1}^M C_L^m.$$

Отсюда найдем коэффициент эмерджентности Хартли  $\varphi$ :

$$\varphi = \frac{\log_2 \sum_{m=1}^M C_L^m}{\log_2 L},$$

который представляет собой относительное превышение количества информации о системе при учете системных эффектов над количеством информации без учета системности. Тем самым коэффициент отражает уровень системности объекта.

Применив полученное значение для коэффициента эмерджентности, получим:

$$I = \log_2 L \frac{\log_2 \sum_{m=1}^M C_L^m}{\log_2 L}.$$

Учитывая, что  $I_{L \rightarrow \infty} \rightarrow L$ , получим:

$$I = \log_2 L^{\frac{L}{\log_2 L}} = L.$$

Следовательно, количество информации в OLAP-кубе равно количеству членов измерения.

Коэффициент эмерджентности Хартли отражает уровень системности объекта и изменяется от 1 (системность минимальна) до  $\frac{L}{\log_2 L}$  (системность максимальна).

#### Выводы

Рассмотрен и предложен достаточно универсальный обновленный подход к моделированию OLAP-кубов, опирающийся на современные онтологические и системные представления в этой области научных знаний.

#### СПИСОК ЛИТЕРАТУРЫ

1. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. Определения, теоремы, формулы / под общей ред. И.Г. Арамановича. – М.: Наука, 1974. – 832 с.
2. Выгодский М.Я. Справочник по элементарной математике. – М.: Физматгиз, 1962. – 420 с.

Поступила 25.01.2010 г.

УДК 004.657

## СЕМАНТИКО-ЭНТРОПИЙНОЕ РЕГУЛИРОВАНИЕ ИНФОРМАЦИОННОГО МОРФИЗМА РЕАЛИЗАЦИЙ xOLAP

А.А. Миронов, А.С. Сигов

Московский государственный институт радиотехники, электроники и автоматики (технический университет)  
E-mail: mironov@mirea.ru

*Анализ опыта создания и сопровождения хранилищ данных говорит о том, что именно в этой области IT индустрии наиболее резко ощущаются трудности, порожденные отсутствием устоявшейся семантической теории информационных процессов и систем. Статья нацелена на изучение моделей xOLAP, целевым образом ориентированных на семантические методы управления, затрагивает понятия семантических разрывов применительно к xOLAP, их семантико-энтропийных оценок и регулирования.*

#### Ключевые слова:

*Оперативная аналитическая обработка данных, семантический разрыв, энтропийное регулирование, информационный морфизм.*

#### Key words:

*On-line analytical processing, semantic break, entropy control, information morphism.*

Разнообразие версий OLAP достаточно велико и расширяется. Модели OLAP обретают новые классификационные признаки, свойства, изменяющие их особенности, достоинства и недостатки, впрочем, оцениваемые в зависимости от специфики решаемых задач. Так, наряду с такими известными модификациями как ROLAP, MOLAP и HOLAP [1], в последние годы появились и находят широкое применение SOLAP (*Spatial On-Line Analytical Processing*) – пространственная аналитиче-

ская обработка, предназначенная для изучения пространственных данных, объединяющая понятия из существенно отличающихся друг от друга сфер знаний, а именно географических информационных систем и OLAP, разработанная для интерактивного и быстрого анализа больших объемов данных; R-ROLAP (*Real-time ROLAP*) – OLAP реального времени, в отличие от ROLAP в R-ROLAP для хранения агрегатов не создаются дополнительные реляционные таблицы, а агрегаты рассчитыва-

ются в момент запроса, многомерный запрос к OLAP-системе автоматически преобразуется в SQL-запрос к реляционным данным [2]; JOLAP – Java OLAP, платформенно-независимый стандарт создания, хранения, доступа и обслуживания данных в OLAP-серверах, основанный на технологии Java, является чистым Java API для J2EE TM, который поддерживает создание и поддержание OLAP данных и метаданных.

В целях более обобщенного и экономного изложения объединим всю известную совокупность разнообразных OLAP единым основообразующим термином – аббревиатурой «xOLAP», понимая под ним различные системы онлайн-аналитической обработки данных в многомерных кубах для поддержки принятия решений (OLAP), где приставка «x» отображает возможное разнообразие модификаций в части особенностей хранения данных, организации запросов к ним и отчетов и т. д. Такого рода обобщение разновидностей OLAP приводит к рекомендации в адрес многочисленных пользователей этих систем ввести обозначение всего класса этих систем в виде аббревиатуры xOLAP, подчеркивая тем самым, с одной стороны их единство на основе извлечения данных из многомерных построений, с другой стороны, расширяющийся спектр их разнообразных функциональных особенностей, обычно определяемый вводимым перед выражением OLAP прилагательным («Multidimensional», «Hybrid» и т. п.), фиксируемым в самом общем виде буквой «x».

Сохранение и углубление семантики обобщения/специализации (*roll-up/drill-down*) становится все более актуальным в обеспечении эффективного моделирования и проектирования xOLAP. Эта парадигма становится не менее важной, чем те, которым уделялось первостепенное внимание в предшествующие годы, а именно, вопросам агрегирования, сортировки, кэширования и группировки данных, сжатия, декомпозиций и реструктуризации кубов и т. д. Даже существенные успехи в решении такого рода задач не снимают остроты вопроса семантики отношений, поскольку по завершению ступени дивергенции проектирования согласно конвергентно/дивергентному методу управления проектами xOLAP восстановление ранее не отрегулированной семантики отношений становится невозможной (даже на ступени трансформации метода, не говоря уже о заключительной ступени конвергенции этого метода), что делает результаты проектирования информационной системы (ИС) неполноценными и тупиковыми.

Более того, возникла и начала удовлетворяться потребность в моделях xOLAP, целевым образом ориентированных на семантические методы управления в статусе главенствующих. Классифицируем такие модели как SeOLAP (Semantic OLAP) и будем настоятельно рекомендовать профессиональному сообществу специалистов этой области придерживаться такого классификационного

признака, поскольку он вполне передает прозрачную аналогию парадигмы SeOLAP со стремительным становлением семантических сетей второго поколения Semantic Web или ONTONET, где также главенствуют принципы семантического регулирования в части поиска и извлечения данных и знаний.

В обеих рассматриваемых ситуациях семантическое управление встает заслоном по отношению к возникновению «взрыва данных» (профессиональный термин теории фактографических информационных систем) и резкому, переходящему в энтропийный, лавинообразному росту энтропии главной функции системы (сети). Вопросы изучения и улучшения моделей семантического управления xOLAP разработаны в специальной литературе недостаточно, хотя в последние годы внимание к ним явно обозначилось. Анализ имеющихся источников в сфере создания и сопровождения хранилищ данных говорит о том, что именно в этой области IT индустрии наиболее резко ощущаются трудности, порожденные отсутствием устоявшейся семантической теории информационных процессов и систем.

Новым в постановке и развитии такого рода раздела теории информационных процессов и систем является то, что парадигмы этого раздела опираются на учет происходящего на интегративной основе слияния инфологий и морфологий архитектур OLAP и Semantic Web второго поколения, то есть ONTONET [3]. Именно на этой основе преодолевается так называемый «семантический разрыв OLAP», о котором пишут в публикациях последних лет. Речь здесь, прежде всего, идет о сборе и консолидации данных из разрозненных и несогласованных источников в предметно-ориентированный, интегрированный и независимый от времени (atemporary system, AtempOLAP) набор данных.

Такое построение уже неизбежно, если возникает опасность коллапсирования фактографической системы по схеме развития «взрыв данных»; оно вообще универсально, если ставится вопрос о перегрузке данных из одного хранилища – донора в другое хранилище – акцептор, созданное независимо от донора. При этом сразу же возникает задача выделения частей и признаков, обладающих абсолютными свойствами мажоритарности и эргодичности, анализируя и соединяя которые можно гармонизировать контент по семантическим признакам. Здесь, согласно отраслевой ВШ РФ «Онтологии ИС» [3], под мажоритарностью функционала ИС понимается одно из важнейших обязательных свойств ИС и сетей, заключающееся в том, что все сигналы, события, команды на входе или в любой части системы или сети согласуются с аналогичными проявлениями на выходе или в других частях сети или системы (кроме специально обособленных); эргодичность рассматривается как явление, при котором средние значения по времени почти всех возможных реализаций процесса с ве-

роятностью единица сходятся к одной и той же постоянной величине; а гармонизация контента является продуктом систематизации и унификации в результате изменения состава, свойств и признаков составляющих контента, приводящих к росту энтропии и мажоритарности системы.

В теории фактографических ИС носителем таких данных, содержащих измеряемые показатели, доступные для хранилища данных, являются OLTP (Online Transaction Processing) системы обработки данных в реальном времени или, что тоже самое, транзакционные системы [4, 5]. Транзакционная система – в информатике, система, реализующая транзакции над хранилищем данных. Задача транзакционной системы – обработать как можно больше транзакций в минимальное время с гарантией безошибочных результатов, то есть пертинентно (релевантности для xOLAP абсолютно недостаточно). Решать такого рода задачу можно только с использованием систем под семантическим дирекционным управлением, то есть используя SeOLAP модельный подход. Этот подход развивает идеологию применения распределенных, многофазных, семантиконесущих и других транзакций в их новых качествах и выстраивая для этого новые модифицированные архитектуры различных версий SeOLAP систем (в настоящую статью предлагаемые авторами альбомы архитектур разновидностей xOLAP и математического описания их информационных морфизмов не внесены ввиду громоздкости изложения, но с ними можно ознакомиться в фондах ОФАП).

Традиционный в теории OLAP подход, связанный с исследованиями образующихся множеств морфем, денотатов (понятий, признаков) вполне универсально пригоден для работы практически со всеми известными авторам разновидностями xOLAP, поэтому здесь подробно не обсуждается. Однако нетрудно заметить, что в разных OLTP-системах одним и тем же понятиям могут оказаться присвоены разные имена и, наоборот, одни и те же имена могут быть присвоены понятиям с разными концептами, то есть возникают пересекающиеся или еще сложнее соотносящиеся друг к другу денотаты, что и вызывает семантический разрыв.

В кибернетическом аспекте последняя задача может эффективно решаться введением в архитектуру xOLAP дирекционных подсистем семантического управления метаданными. Дирекционная семантическая подсистема, видимо, должна проектироваться с учетом того, что вся система SeOLAP в целом должна иметь общий репозиторий, в котором хранятся как семантические слои, так и права пользователей (логинов) на объекты семантических слоев, при этом типы семантических слоев зависят от типов используемых источников данных. Для создания и поддержки семантических слоев используются различные приложения, число которых на рынке IT быстро увеличивается (в статье обзор опущен).

На основе преодоления семантических разрывов всех возможных разновидностей и их комбинаций, в сущности, осуществляется интеграция xOLAP и Semantic Web с образованием SeOLAP. По мнению некоторых авторитетных авторов [6] единение Semantic Web – скорее парадигма, чем исчерпывающее решение вопроса о снятии проблемы семантических разрывов xOLAP. Действительно, сегодняшнее развитие Semantic Web сконцентрировано главным образом в направлении дополнительной метаинформации к документу, чем на семантической интеграции самих данных. Появившийся недавно новый вариант развития Fusionsoft Semantic Net преодолевает это затруднение, поскольку он ориентирован на семантическую интеграцию данных [7]. К тому же обеспечивается автоматическая навигация по признакам дополнительно вводимой семантически связанной информации, причем по гетерогенным и распределенным признакам без необходимости писать программный код. Следует заметить при этом, что платформы Fusionsoft Semantic Net и Web дополняют друг друга. Развивая линию на создание и достаточно широкое использование отраслевой унифицированной классификации ИС авторы настоящей публикации предлагают обозначать описанную выше модификацию как FSeOLAP (Fusionsoft Semantic OLAP), относя к этому названию результат инфологического строительства с использованием Fusionsoft Semantic Net.

Еще один пример интегративного строительства xOLAP с применением семантических принципов управления – TransSeOLAP или Semantic OLAP Transformer на основе использования технологии Panorama.

С позиций дальнейшего углубления теории информационных процессов и систем в моделировании функционала всех этих систем присутствует единый принцип изучения и упорядочения информационных морфизмов как системы с пользователем, так и на подсистемных уровнях и уровнях слияния и взаимодействия на единых семантических принципах различных баз данных с различных серверов, да еще возможно в интенсивном многопоточном запараллеленном режиме. Соответственно, в математические модели информационных морфизмов всех этих уровней и комбинаций взаимодействий должны органически вписываться в качестве основоопределяющих составляющие семантико-энтропийного регулирования.

Информационный морфизм представляет собой класс эквивалентности, взаимодействие. Информационный морфизм интерпретируется здесь как гомоморфизм свободного моноида в информационном поле, генерируемого из сообщества морфологических, иногда и синтаксических, схожеств и признаков, способных к кластеризации, что принципиально важно в условиях главенствования семантических признаков и принципов управления.

Возникающие носители, в частности, упомянутые выше семантические слои xOLAP, могут обладать или не обладать устойчивостью по отношению к информационной среде. При появлении устойчивого носителя может происходить фиксация возникшего типа носителя в случае возможного его использования по отношению к информационной структуре более высокого порядка. Отсюда вероятностная модель информационного морфизма  $V$  между двумя подсистемами, слоями или отображениями взаимодействий (например, в формировании отчетов)  $A$  и  $B$  в информационной среде определяется следующим образом:

$$V_i = C_i / E_a + k * E_b, \quad (*)$$

где  $C_i$  – относительное количество информации вида  $I$  в дуплексном (самый общий случай информационного обмена между объектами  $A$  и  $B$ ) информационном пространстве;  $E_a$  и  $E_b$  – относительные (долевые) распределения информации в потоках в направлениях от  $A$  к  $B$  и от  $B$  к  $A$ ;  $k$  – сложный коэффициент, в первом приближении равный натуральному числу  $e$  в степени произведения:  $-L(G_{ai}-G_{bi})$ , где  $L$  – коэффициент Лагранжа,  $G_{ai}$  и  $G_{bi}$  – характеристические коэффициенты информационных потоков в направлениях от  $A$  к  $B$  и от  $B$  к  $A$ .

Модель (\*) позволяет отследить основные закономерности информационного морфизма. Показателем упорядоченности в модели является информационная энтропия взаимодействующих объектов, что является классикой семантико-энтропийных оценок и регулирования.

В самом общем виде семантико-энтропийные оценки и регуляторы используют понятие обобщенной энтропии, которое в первоуровневом ее виде (опуская промежуточные выкладки и рассуждения) найдено авторами настоящей статьи слишком общим и неточно отображающим все внутрисистемные взаимодействия SeOLAP и модификаций, поскольку возникновение семантических разрывов провоцируется не только сугубо морфологическими причинами, но и проблемами многопоточности, кроссязыковыми и другими, описанными выше.

Найдено, что в отношении основной версии модельного представления SeOLAP достаточно универсален и продуктивен семантико-энтропийный анализ с использованием комплексной реализации так называемых условной энтропии, взаимной энтропии и энтропии объединения (впрочем, не лишено интереса применение других разновидностей, например, энтропии потока, кросс-энтропии и других).

Остановимся на этом вопросе подробнее. Выше упоминалось, что в моделях всех xOLAP отклик всегда должен быть пертинентным (релевантным он уж точно является, но не наоборот). В этой си-

туации лучше воспользоваться не энтропией информации, а так называемой условной энтропией, в названии которой озвучена некая назначенная пользователем условная зависимость вероятностей различных событий друг от друга. Вид этой зависимости определяет форму математического описания такой энтропии, которая может оказаться весьма сложной. Разнообразия здесь много. Поэтому здесь приводится наиболее очевидный случай такой зависимости, аналогичный марковской модели первого порядка. Последовательность дискретных случайных величин  $\{X_n\}_{n \geq 0}$  называется цепью Маркова (с дискретным временем, поскольку априори рассматривается aTemoгу), если:

$$P(X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = i_{n+1} | X_n = i_n).$$

Здесь в простейшем случае условное распределение последующего состояния цепи Маркова зависит только от текущего состояния и не зависит от всех предыдущих состояний (в отличие от цепей Маркова высших порядков). По аналогии условной энтропией первого порядка (филогенетической информативностью) является энтропия для алфавита, где известны вероятности появления одной буквы после другой (т. е. вероятности двухбуквенных сочетаний) или определена вероятность и даже неизбежность возникновения одного события, свойства, явления за другим (например, неизбежность принятия решения о релевантности отклика, если он пертинентен). Для двухбуквенной (двухуровневой) зависимости первого порядка условная энтропия в самом простейшем случае может иметь следующее математическое описание:

$$H_1(S) = - \sum_i p_i \sum_j p_i(j) \log_2 p_i(j).$$

где  $i$  – состояние, зависящее от предшествующего символа, и  $p_i(j)$  – это вероятность  $j$ , при условии, что  $i$  был предыдущим символом.

Через частную и обобщенную условные энтропии полностью описываются информационные семантические сбои и риски (назовем их для краткости последующего изложения aStvantic Agent – aSA) при передаче данных и любой транзакции, условно считая ее канальной. Для этого в предметной области «телематика» используют понятие так называемых канальных матриц. Воспользуемся им – опять же прибегая к методу аналогий. Так, в телематике для описания потерь со стороны источника (известен посланный сигнал), рассматривается условная вероятность  $p(b_j|a_i)$  получения приёмником символа  $b_j$  при условии, что был отправлен символ  $a_i$ . В рассматриваемом случае замена (риск замены) истинного значения и понимания  $a_i$  на  $b_j$  и есть проявление семантического разрыва в xOLAP. При этом матрица совокупности значений aSA по аналогии с теорией телематики обретает следующий вид:

	$b_1$	$b_2$	...	$b_j$	...	$b_m$
$a_1$	$p(b_1 a_1)$	$p(b_2 a_1)$	...	$p(b_j a_1)$	...	$p(b_m a_1)$
$a_2$	$p(b_1 a_2)$	$p(b_2 a_2)$	...	$p(b_j a_2)$	...	$p(b_m a_2)$
...	...	...	...	...	...	...
$a_i$	$p(b_1 a_i)$	$p(b_2 a_i)$	...	$p(b_j a_i)$	...	$p(b_m a_i)$
...	...	...	...	...	...	...
$a_m$	$p(b_1 a_m)$	$p(b_2 a_m)$	...	$p(b_j a_m)$	...	$p(b_m a_m)$

Вероятности, расположенные по диагонали описывают вероятность истинного исхода транзакций, а сумма всех элементов столбца даст вероятность появления соответствующего символа на принимающей стороне –  $p(b_j)$ . Тогда семантические риски, приходящиеся на  $a_i$ , описываются через частную условную энтропию вида:

$$H(B | a_i) = - \sum_{j=1}^m p(b_j | a_i) \log_2 p(b_j | a_i).$$

Для вычисления совокупного риска от всех выявленных или возможных аSA может использоваться обобщенная условная энтропия:

$$H(B | A) = \sum_i p(a_i) H(B | a_i).$$

$H(B|A)$  означает энтропию со стороны источника, аналогично рассматривается  $H(A|B)$  – энтропия с принимающей стороны: вместо  $p(b_j|a_i)$  всюду указывается  $p(a_i|b_j)$  (суммируя элементы строки можно получить  $p(a_i)$ , а элементы диагонали означают вероятность того, что был отправлен именно тот семантический посыл, который получен, то есть вероятность отсутствия риска семантического разрыва).

Взаимосвязь переданных и полученных данных в самом общем виде описывается вероятностями совместных событий  $p(a_i b_j)$ , и для полного описания характеристик этого обобщенного процесса требуется только одна матрица вида:

$p(a_1 b_1)$	$p(a_1 b_2)$	...	$p(a_1 b_j)$	...	$p(a_1 b_m)$
$p(a_2 b_1)$	$p(a_2 b_2)$	...	$p(a_2 b_j)$	...	$p(a_2 b_m)$
...	...	...	...	...	...
$p(a_i b_1)$	$p(a_i b_2)$	...	$p(a_i b_j)$	...	$p(a_i b_m)$
...	...	...	...	...	...
$p(a_m b_1)$	$p(a_m b_2)$	...	$p(a_m b_j)$	...	$p(a_m b_m)$

Для более конкретного случая, когда исследуется информационный морфизм взаимодействия двух подсистем исключительно на семантическом уровне, матрица необязательно должна быть квадратной. Очевидно, сумма всех элементов столбца с номером  $j$  даст  $p(b_j)$ , сумма строки с номером  $i$  есть  $p(a_i)$ , а сумма всех элементов матрицы равна 1. Совместная вероятность  $p(a_i b_j)$  событий  $a_i$  и  $b_j$  вычисляется как произведение исходной и условной вероятности

$$p(a_i b_j) = p(a_i) p(b_j | a_i) = p(b_j) p(a_i | b_j).$$

Как показал первоначальный опыт работы с развернутым выше математическим описанием, вполне универсально условные вероятности оцениваются по формуле Байеса. Теорема Байеса – одна из основных теорем элементарной теории вероятностей, которая определяет вероятность наступления события в условиях, когда на основе наблюдений известна лишь некоторая частичная информация о событиях. По формуле Байеса можно более точно пересчитывать вероятность, беря в учёт как ранее известную информацию, так и данные новых наблюдений.

Формула Байеса имеет следующий вид:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)},$$

где  $P(A)$  – априорная вероятность гипотезы  $A$  (смысл такой терминологии см. ниже);  $P(A|B)$  – вероятность гипотезы  $A$  при наступлении события  $B$  (апостериорная вероятность);  $P(B|A)$  – вероятность наступления события  $B$  при истинности гипотезы  $A$ ;  $P(B)$  – вероятность наступления события  $B$ .

Формула Байеса позволяет «переставить причину и следствие»: по известному факту события вычислить вероятность того, что оно было вызвано данной причиной. Это свойство весьма привлекательно при исследовании морфизмов xOLAP, для которых реализуются алгоритмы вычисления кубов по методам обратных запросов (*Iceberg* – кубы). События, отражающие действие «причин», в этом случае называют гипотезами, так как они – предполагаемые события, повлекшие это событие. Безусловную вероятность справедливости гипотезы называют априорной (насколько вероятна причина вообще), а условную – с учетом факта произошедшего события – апостериорной (насколько вероятна причина оказалась с учетом данных о событии). Важным следствием формулы Байеса является формула полной вероятности события, зависящего от нескольких несовместных гипотез, которые иногда могут иметь место для модельных описаний объединяемых в единую систему различных платформ и технологий (как это показано выше целым рядом примеров).

$$P(B) = \sum_{i=1}^N P(A_i) P(B | A_i) - \text{вероятность наступления события } B,$$

зависящего от ряда гипотез  $A_i$ , если известны степени достоверности этих гипотез (например, измерены экспериментально);

Таким образом определяются все данные для вычисления энтропий передающей и принимающей стороны:

$$H(A) = - \sum_i \left( \sum_j p(a_i b_j) \log \sum_j p(a_i b_j) \right),$$

$$H(B) = - \sum_j \left( \sum_i p(a_i b_j) \log \sum_i p(a_i b_j) \right).$$

Взаимная энтропия вычисляется последовательным суммированием по строкам (или по столбцам) всех вероятностей матрицы, умноженных на их логарифм:

$$H(AB) = -\sum_i \sum_j p(a_i b_j) \log p(a_i b_j).$$

Пути несложных преобразований также получаем

$$H(AB) = H(A) + H(B | A) = H(B) + H(A | B).$$

#### СПИСОК ЛИТЕРАТУРЫ

1. Thomsen E. OLAP Solutions: Building Multidimensional Information Systems Second Edition. Wiley Computer Publishing John Wiley & Sons, Inc., 2002.
2. Laks V.S. Lakshmanan, Jian Peiz, Yan Zhao. Qctrees: An efficient summary structure for semantic OLAP. In: SIGMOD, 2003.
3. Мордвинов В.А. Онтология моделирования и проектирования семантических информационных систем и порталов: Справочное пособие. – М.: МИРЭА, 2005. – 237 с.
4. Laks V.S. Lakshmanan, Jian Peiz, Yan Zhao. Socqet: Semantic OLAP with compressed cube and summarization. In: SIGMOD, 2003.
5. Xiaolei Li, Dong Xin Jiawei, Benjamin W. Wah. Star-cubing: Computing iceberg cubes by top-down and bottom-up integration. In: VLDB, 2003.
6. Иванников А.Д., Кулагин В.П., Мордвинов В.А., Найханова Л.В., Овезов Б.Б., Тихонов А.Н., Цветков В.Я. Получение знаний для формирования информационных образовательных ресурсов. – М.: ФГУ ГНИИ ИТТ «Информика», 2008. – 440 с.
7. Yan Zhao. Quotient Cube and QC-Tree: Efficient Summarizations for Semantic OLAP. 2003.

Поступила 25.01.2010 г.

УДК 004.822

## ВОПРОСЫ МАТЕМАТИЧЕСКОГО ОБЕСПЕЧЕНИЯ КОМПЬЮТЕРНЫХ СЕТЕЙ НА ОСНОВЕ СЕМАНТИЧЕСКОГО ПОДХОДА

А.Ю. Войтович, В.А. Мордвинов

Московский государственный институт радиотехники, электроники и автоматики (технический университет)  
E-mail: voytovich@mirea.ru

*Рассматривается обновление подходов к моделированию и формализации описания информационного морфизма на межслойных переходах семантической структуры при сохранении гарантированных показателей pertinентности, релевантности и когнитивности информационных систем, построенных на основе расслоенных архитектур.*

#### Ключевые слова:

*Семантическая сеть, информационный морфизм, «сэндвич» Бернерса-Ли, весовой коэффициент, межуровневые переходы, позиции семантико-энтропийное регулирование.*

#### Key words:

*Semantic network, information morphism, «sandwich» of Berners-Lee, weight coefficient, interlevel transitions, semantic-entropy control.*

Согласно концепции Тима Бернерса-Ли семантическая сеть является интегральным понятием, объединяющим ряд наиболее перспективных направлений развития Интернета и представляет собой многослойную архитектуру, каждый последующий уровень которой отвечает за более тонкие механизмы представления и обработки данных и знаний [1]. Эта структура, рисунок, получила в научных источниках название «сэндвича» Бернерса-Ли:

В основе концепции лежат стандарты, спецификации и рекомендации консорциума W3C (World Wide Web Consortium) для расширяемого

Взаимная энтропия обладает свойством информационной полноты – из неё можно получить все рассматриваемые величины.

Достижимые в результате моделирования полнота вероятностей событий и информационная полнота дают основания применять подходы и результаты показанного здесь моделирования к достаточно широкому спектру разновидностей хО-OLAP, в том числе FSeOLAP, TransSeOLAP и другим, представленным в альбоме классификаций OLAP.

языка разметки XML, языка описания структуры XML документа XML Schema, языка запросов XQuery, словаря RDF, унифицированного идентификатора ресурса URI, языка онтологии сетевых сервисов OWL-S [3]. Прогресс концепции заключается в том, чтобы последовательно стандартизировать «семантический сэндвич» – слой за слоем, снизу вверх. Таким образом, выстраиваются механизмы функционирования на каждом уровне. Необходимые средства для формализации информационных процессов на каждом отдельно взятом уровне уже выработаны – их в должном количестве