

УДК 004.318

А.А. МОРОЗОВ, В.И. ВЬЮН, Г.Е. КУЗЬМЕНКО

**ПРОБЛЕМНО-ОРИЕНТИРОВАННЫЙ КОНТЕНТ-АНАЛИЗ
В СТРУКТУРЕ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ (СППР)**

Abstract: It is discussed one from approaches for providing viable development. SSMD on the base of using interactive methods of intelligent analysis of data of the past activity <SSMD+Object>. The corresponding set of mechanisms and instruments that broaden intelligent ability of the traditional SSMD is proposed to consider as problem-oriented content analysis of "history" activity of association.

Key words: content-analysis, SSMD, ontology.

Анотація: Розглядається один із підходів щодо забезпечення життєздатного розвитку СППР на основі використання інтерактивних методів інтелектуального аналізу даних минулої діяльності об'єднання <СППР+ОБ'ЄКТ>. Відповідний набір механізмів та інструментів, що розширюють "інтелектуальні" здібності традиційної СППР, пропонується розглядати як проблемно-орієнтований контент-аналіз "історії" діяльності об'єднання.

Ключові слова: контент-аналіз, СППР, онтологія.

Аннотация: Рассматривается один из подходов к обеспечению жизнеспособного развития СППР на основе использования интерактивных методов интеллектуального анализа прошлой деятельности объединения <СППР+ОБЪЕКТ>. Соответствующий набор механизмов и инструментов, расширяющих "интеллектуальные" способности традиционной СППР, предлагается рассматривать как проблемно-ориентированный контент-анализ "истории" деятельности объединения.

Ключевые слова: контент-анализ, СППР, онтология.

1. Введение

Одним из аспектов искусственного интеллекта является то, что это самообучающийся инструмент, усиливающий мыслительную деятельность человека по генерации и принятию решений [1]. Исходя из этого, повышение уровня интеллекта системы поддержки принятия решений (СППР) можно рассматривать как экспериментальную научную дисциплину, в которой роль эксперимента выполняет "история" фактической деятельности системы как информационная платформа процессов самообучения и саморазвития.

В этом плане под интеллектуализацией СППР будем понимать расширение традиционной структурно-функциональной организации (базовой) СППР проблемно-ориентированными механизмами и инструментами содержательного анализа результатов фактической деятельности объединения <СППР+ОБЪЕКТ> по сценарию {вчера/сегодня → сегодня/завтра}.

Цель анализа – выявление тенденций и неизвестных ранее закономерностей функционирования <СППР+ОБЪЕКТ> для формирования, а при необходимости и упреждающего внесения релевантных изменений в структурно-функциональную организацию базовой СППР.

В работах [2 – 4] были предложены системотехнические механизмы поддержки анализа результатов деятельности <СППР+ОБЪЕКТ> на базе Хранилища Данных (Хрд). В данной работе предлагается рассмотреть содержательные аспекты операций и процедур динамического ведения Хрд как "истории" функционирования объединения.

2. Постановка задачи

2.1. Одинаковое (аналитиком и системой) восприятие результатов деятельности, поступающих в ХрД, в определенной степени достигается заданием открытого множества ключевых параметров-показателей (КПП) деятельности <СППР+ОБЪЕКТ>, однозначная интерпретация которых обеспечивается на базе априорной онтологии [5] <СППР+ОБЪЕКТ> и КПП.

В данном случае под заданием КПП мы имеем в виду представление профессиональных знаний и навыков специалистов аналитических служб ОБЪЕКТА в операциях содержательной обработки оперативных данных его деятельности.

Эти представления и определяют начальную, или априорную, онтологию <СППР+ОБЪЕКТ>. Оценка результатов содержательного анализа входного потока оперативных данных деятельности на платформе априорной онтологии является для аналитика источником возможных идей и гипотез по обоснованию необходимости ее модернизации или построения более релевантной текущим условиям и требованиям к функционированию СППР&ОБЪЕКТ апостериорной модели онтологии.

Отметим, что именно итерационный процесс модернизации/адаптации моделей онтологии и является объективным отражением процесса эволюционного развития “интеллекта” <СППР+ОБЪЕКТ>.

2.2. Почему могут не выполняться принятые решения? Что такое “правильные” решения и почему со временем меняются критерии “правильности”? Как такие изменения отражаются на функциональной структуре базовой СППР?

Так как результаты текущего функционирования объединения оценивает человек, то всегда будет возникать проблема адекватной их интерпретации, а, значит, факты несоответствия ожидаемому будут проявляться вновь и вновь [6].

Имеет ли право аналитик, обнаруживший определенную закономерность в проявлении несоответствия и несогласованности фактических и ожидаемых значений КПП по результатам наблюдения за деятельностью <СППР+ОБЪЕКТ>, безапелляционно квалифицировать их как ошибки в работе базовой СППР? По нашему мнению, интерпретацию подобных событий типа – это проявление ранее неизвестной семантической и/или логической не следует игнорировать. Связанная с ней возможность осознания необходимости уточняющих изменений в модели онтологии и технологических процессах функционирования делает систему более жизнеспособной.

Естественно, с целью избежать возможного “эффекта маятника” при внесении изменений в структурно-функциональную организацию <СППР+ОБЪЕКТ> следует учитывать следующее: реализации подлежат лишь изменения, полученные на основе длительного анализа и тестирования устойчивости новых зависимостей между КПП онтологических моделей объединения.

Онтологии и онтологическое моделирование в настоящее время представляют весьма популярное, быстро развивающееся (и вширь, и вглубь) научно-техническое направление системных исследований конкретных динамически развивающихся объектов.

Будем считать, что онтология <СППР+ОБЪЕКТ> задает внешнюю аппроксимацию (“горизонты” основных аспектов существования – цель, назначение, ресурсные атрибуты (материальные и нематериальные) деятельности, их взаимоотношения и взаимодействия и др.)

представления предметной области (ПрО) [5]. Как всегда, практика опережает теоретическое осмысление проблемы. Границы и градации используемых понятий со временем оказываются размытыми, и в какой-то момент участники процесса начинают испытывать потребность в том, чтобы переопределить концептуальные рамки проекта.

2.3. Структура входного потока атрибутов {ОБД→Хрд} является неоднородной совокупностью информационных объектов. Это записи реляционных БД, текстовые файлы и сообщения (включая внешние источники), графики и таблицы, документы и т.п. Вследствие этого, предлагается пользоваться более широким пониманием термина контент-анализ, определяя его как проблемно-ориентированный процесс на добывание (data mining) новых знаний.

В данной работе термин контент-анализ охватывает все типы содержательной обработки информационных объектов входного потока, включая всевозможные семантические зависимости и отношения между ними типа подобных по структуре поведения, имеют одинаковое содержание и т.п. В этом контексте расширение базовой СППР непосредственно зависит от операционного обеспечения процессов проблемно-ориентированного контент-анализа (ПОКАН) с применением онтологических описаний <СППР+ОБЪЕКТ> и КПП.

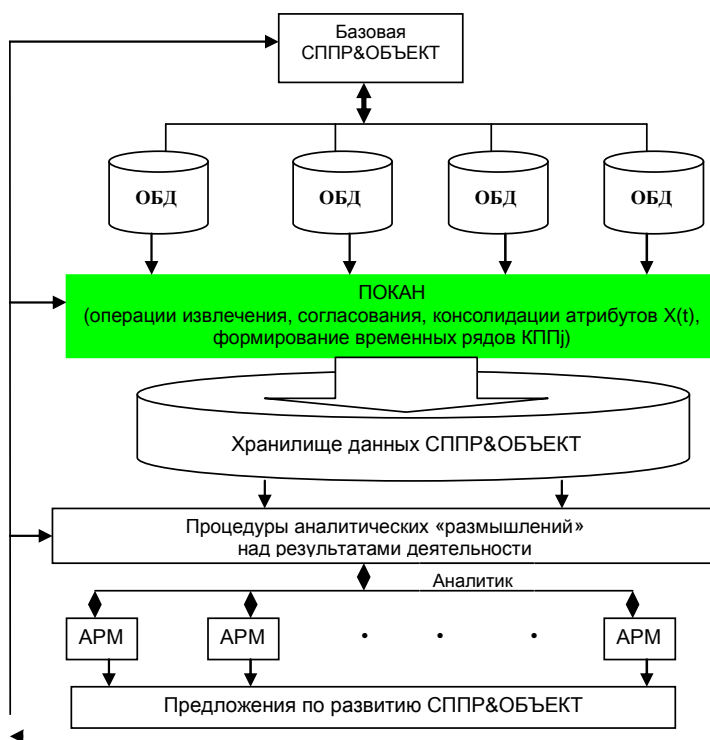


Рис. 1. Схема взаимодействия компонент процесса интеллектуализации СППР на основе ПОКАН

3. Базисные технологии ПОКАН

Схема взаимодействия всех компонент процесса повышения уровня интеллекта СППР &ОБЪЕКТ на основе содержательного анализа результатов его фактической деятельности представлена на рис. 1.

Рассмотрим функции блока ПОКАН, который выполняет посредническое или пороговое взаимодействие базовой СППР и ее интеллектуального расширения на платформе Хрд.

С нашей точки зрения, для порогового взаимодействия важным моментом является фиксация текущих значений составляющих КПП атрибутов

(оперативных данных деятельности) в виде многомерного вектора-выборки из множества $X(t)$ атрибутов ОБЪЕКТА, распределенного по некоторой совокупности оперативных баз данных (ОБД) базовой СППР. Векторы-выборки отображают текущие значения КПП, которые должны быть доступными для модулей последующих контент-операций. Таким образом, исполнительные

подсистемы СППР заканчивают очередной этап выполнения принятых решений – формирование совокупности зафиксированных t -значений атрибутов $X(t)$, которые и рассматриваются как входной поток порогового взаимодействия {ОБД=>ХрД}.

В работах [5, 6] определены основные контент-операции обработки входного потока данных от ОБД для размещения их в специализированное ХрД. Это операции извлечения, согласования, консолидации атрибутов $X(t)$, формирование временных рядов КПП.

Входной поток оперативных данных для ПОКАН является потоком копий текущих значений атрибутов $X(t)$. Начальная операция контент-анализа – извлечение оперативных данных из ОБД – поддерживается интерфейсными модулями блока, адаптированными к оболочкам ОБД и действующей модели онтологии <СППР+ОБЪЕКТ>. Последующие контент-операции будем рассматривать как “разворачивание панорамы” целостного представления СППР&ОБЪЕКТ в виде многомерного отображения t -вектора атрибутов $X(t) = \{a_i(t) | i = 1, 2, \dots, M\}$, где M – общее количество атрибутов.

К ним относятся следующие операции:

- выборки атрибутов и формирование на данный момент времени текущих значений многомерного пространства $\{KПП_j | j = 1, 2, \dots, k\}$ как подмножеств $X_j(t) \in X(t)$ в виде временного ряда (ВР) $X_j(t) = \{a_i^j(t_s) | i = 1, 2, \dots, m_j; s = 1, 2, \dots, n\}$. Отметим, что некоторые атрибуты могут входить в состав нескольких КПП; и, следовательно, $m_1 + m_2 + \dots + m_k \geq M$;

- сглаживание рядов $X_j(t_i)$ с целью устранения случайных всплесков значений, а также временного согласования полученных данных, заполняя “пропуски” интерполяционными значениями;

- выявление поведенческих характеристик поведения ВР:

- * тренд – отклонения значений $\overline{a_i(t)}$ сглаженного ряда от исходного значения $a_i(t)$ на данном временном интервале. Формально представляется в виде трендовых кривых, обычно линейных или нелинейных. Используется в прогнозных оценках “сегодня/завтра”;

- * поворотные точки – точки локальных экстремумов типа $\{a_{i-1} < a_i > a_{i+1}\}$ и $\{a_{i-1} > a_i < a_{i+1}\}$. Используются в оценках “подобия” поведения ВР атрибутов и КПП;

- * фазы ВР (интервалы монотонности) – отрезки ряда между двумя поворотными точками;

- формирование Витрин Данных информационного отображения конкретного КПП_j в составе ХрД;

- завершение операций порогового взаимодействия – прогнозирование “сегодня-завтра”, выявление неизвестных ранее тенденций, закономерностей и взаимозависимостей в поведении КПП “вчера-сегодня” – выполняется совместно с аналитиком в процессах интерактивного взаимодействия, инициируемых как аналитиком, так и системой ведения ХрД по результатам “знакомства с последними поступлениями” входного потока.

4. Пример

Рассмотрим предложенную технологию ПОКАН на примере.

1) Онтология ОБЪЕКТА (рис. 2):

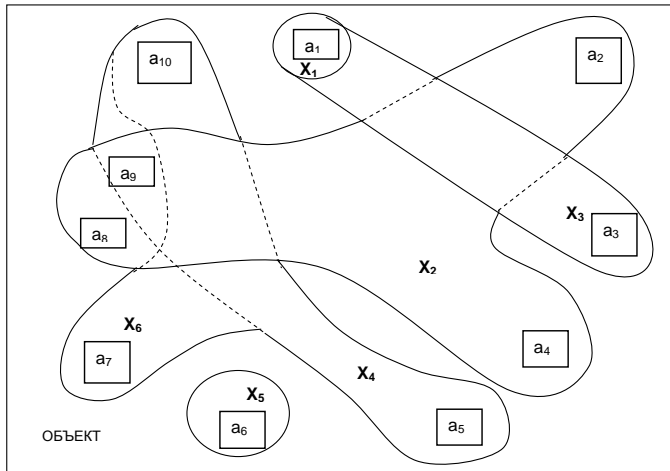


Рис. 2. Онтология ОБЪЕКТА

– перечень атрибутов состояния ОБЪЕКТА –

$$X(t) = \{a_i(t) \mid i = 1, 2, \dots, 10\};$$

– описание КПП ОБЪЕКТА (табл. 1);

– описание отношений и зависимостей между КПП:

$$X_1(t) \cap X_3(t) = X_1(t),$$

$$X_2(t) \cap X_6(t) = \{a_9\}.$$

Таблица 1. Описание КПП объекта

X_j	Атрибуты	m_j
X_1	$\{a_1^1\}$	1
X_2	$\{a_2^2, a_4^2, a_8^2, a_9^2\}$	4
X_3	$\{a_3^3, a_1^3\}$	2
X_4	$\{a_5^4, a_9^4, a_{10}^4\}$	3
X_5	$\{a_6^5\}$	1
X_6	$\{a_7^6, a_5^6, a_{10}^{6j}\}$	3

2) Хранилище данных (табл. 2).

Таблица 2. Отображение n -регистраций совокупности t -значений атрибутов

$a_1(t_1)$	$a_2(t_1)$	$a_3(t_1)$	$a_4(t_1)$	$a_5(t_1)$	$a_6(t_1)$	$a_7(t_1)$	$a_8(t_1)$	$a_9(t_1)$	$a_{10}(t_1)$
$a_1(t_2)$	$a_2(t_2)$	$a_3(t_2)$	$a_4(t_2)$	$a_5(t_2)$	$a_6(t_2)$	$a_7(t_2)$	$a_8(t_2)$	$a_9(t_2)$	$a_{10}(t_2)$
...
$a_1(t_n)$	$a_2(t_n)$	$a_3(t_n)$	$a_4(t_n)$	$a_5(t_n)$	$a_6(t_n)$	$a_7(t_n)$	$a_8(t_n)$	$a_9(t_n)$	$a_{10}(t_n)$

3) Формирование атрибутного состава, например, $\{X_3\}$ как витрины Хрд.

Таблица 3. Атрибутный состав $\{X_3\}$

Атрибут a_3 (сглаженный ряд)	Тренд (лин./нелин.)	Поворот. точки (ai(tj))	Фаза монотон. $[t_k, t_1]$	Атрибут a_1 (сглажен. ряд)	Поворот. точки (ai(tj))	Фаза монотон. $[t_k, t_1]$	Тренд (лин./нелин.)
$\overline{a_3}(t_1)$	л	$a_3(t_s)$	$[t_1, t_s]$	$\overline{a_1}(t_1)$	---	$[t_1, t_n]$	л
$\overline{a_3}(t_2)$				$\overline{a_1}(t_2)$			
...	нл	---	$[t_1, t_n]$	л
$\overline{a_3}(t_n)$				$\overline{a_1}(t_n)$			

Анализ примера.

Приведенный фрагмент работы ПОКАН по обработке данных входного потока отражает работу интеллектуальных модулей по формированию ХрД и витрины ХрД для КПП X_3 на платформе действующей онтологии ОБЪЕКТА (рис. 2).

Практическую пользу, которую представляют полученные результаты, сформулируем в виде следующих методологических положений [7], определяющих сущность практического интеллекта (умение адаптироваться к окружающей обстановке, добиваясь поставленных целей) ПОКАН:

- формирование Витрин ХрД по отображению реального состояния КПП ($X_1 - X_6$);
- обеспечение более полного восприятия аналитиком характеристик реального поведения объединения в изменяющихся условиях внешней среды;
- разработка альтернативных направлений адаптации системы к новым условиям;
- формирование релевантных к новым условиям существования изменений в онтологической модели <СППР+ОБЪЕКТ> и КПП.

Отметим также, что в примере не отражены результаты контент-операций, связанных с выявлением неизвестных ранее зависимостей между КПП, или их взаимовлияние, которые несомненно обогащают “аналитические размышления”.

Вместе с этим аналитикам должна быть предоставлена возможность непосредственного доступа к ХрД.

5. Выводы

В работе введено понятие проблемно-ориентированного контент-анализа (ПОКАН) как процесса, направленного на добывание новых знаний из фактов реальной деятельности объединения <СППР+ОБЪЕКТ>. Сформулированы методические положения реализации блока ПОКАН на базе онтологической модели объединения.

Предложен общесистемный набор контент-операций поддержки порогового взаимодействия “базовая СППР \leftrightarrow система формирования ХрД”.

По нашему мнению, этот подход обеспечивает создание активной информационной платформы для “аналитических размышлений” экспертов-аналитиков над проблемами деятельности и, при необходимости, разработки предложений по изменениям в структурно-функциональной организации <СППР+ОБЪЕКТ>.

“Базовый хребет” предложения определяется простым и понятным условием – онтология любого объединения <ИС+ОБЪЕКТ> (не только СППР) должна иметь внутреннюю функциональность, основу которой составляют модули контент-операций целостного представления объединения.

СПИСОК ЛИТЕРАТУРЫ

1. Шемакин Ю.И. Системантика: (Монография). – М.: Изд-во РАГС, 2006. – 266 с.
2. В'юн В.І., Кузьменко Г.Є. Інтелектуалізація ІС: від інформаційно-пошукових тезаурусів до онтологічних схем змістового аналізу результатів діяльності системи // Доклади ІХ междунар. научн. конф. ім. Т.А. Таран “Інтелектуальний аналіз даних” ІАІ-2009. – Київ, 2009. – 19–22 мая. – С. 68 – 72.

3. Морозов А.О., В'юн В.І., Кузьменко Г.Є. Інтелектуалізація інформаційних систем: орієнтація на формування знань в процесах аналізу "інформаційних згорток" // Математичні машини та системи. – 2005. – № 2. – С. 140 – 146.
4. Асельдеров З.М., В'юн В.І., Морозов А.О. Континуум розумності інформаційних систем // Матеріали конф. «Искусственный интеллект». – Донецьк, Кацивелі, 2004. – Т. 2. – С. 167 – 168.
5. Артемьева И.Л., Гаврилова Т.Л., Клещев А.С. Модели предметной области с атомарными объектами // НТИ. – 2004. – № 12. – С. 8 – 17.
6. Балабанов А.С. Выделение знаний из баз данных – передовые компьютерные технологии интеллектуального анализа данных // Математичні машини і системи. – 2001. – № 1–2. – С. 40 – 54.
7. Мейтус В.Ю. Программирование и интеллект // Проблемы программирования. – 2009. – № 1. – С. 17 – 26.

Стаття надійшла до редакції 18.03.2009