

Вертикаль поддержки принятия решений или почему 2+2 не всегда равно 4

СППР – системы поддержки принятия решений – это класс ИТ-систем, ориентированных на поддержку менеджмента при выработке им решений. О некоторых проблемах при создании такой системы и примере ее реализации идет речь в данной статье

Поддержка принятия решений

Пирамида источников. Рассмотрим гипотетическую разведку. Десять источников поставляют информацию агенту, десять агентов поставляют информацию разведчику, десять разведчиков передают информацию резиденту, десять резидентов отправляют разведанные в центр, аналитик генерального штаба получает сводки из десяти центров разведок разных родов войск и направлений, разработчик военных, дипломатических и прочих операций получает обзоры ситуации от десяти аналитиков, верховный командующий получает десять вариантов планов по проведению военной, политической или другой операции, анализирует положительные и отрицательные последствия планируемых действий (прогнозируемые), и на основе общего видения ситуации принимает решение об осуществлении того или иного варианта плана. Далее еще следует разбор результатов принятого решения и т.д. Вот это и есть то, что называют «вертикалью поддержки принятия решения». См. рис. 1. Мы не будем рассматривать действия выше уровня принятия решений, для примера хватит и первых пяти.

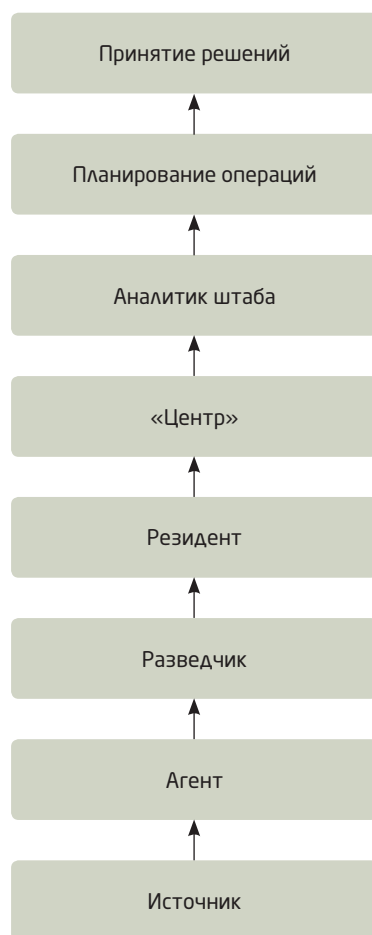


Рис. 1. Вертикаль продвижения информации

В приведенном простом примере вертикали продвижения информации не учитывается ряд моментов практической работы¹, например, таких как присвоение коэффициента важности информации (иногда информация «у него хлопнула форточка» куда важнее вопля «завтра будет война!»), проверки ее достоверности и т.д. – мы же рассматриваем не настоящую разведку, это просто модель.

Но в любом случае в пирамиде есть четкое ранжирование «первичный документ – статистика – аналитика – планирование (прогнозирование) – принятие решений». При этом на каждом уровне обработки информации в нашей модели нагрузка составляет 10 источников и один потребитель информации. Конечно, в реальной разведке источников может быть чуть больше или меньше.

Общий пул. Теперь посмотрим на ту практику, которая сейчас превалирует в подходах к реализации систем поддержки принятия решений (практически все системы служат, прежде всего, для принятия решений, даже прогноз погоды). Человечество обзавелось компьютерами, и кто-то предположил: а что если применять не вертикаль при-

¹См. практические советы по анализу данных на основе опыта работы разведслужб в статьях «Достоверность и интерпретация данных» ИТМ № 4/2010, «7 этапов информационно-аналитической работы» ИТМ № 5/2010, «5 факторов успешного предвидения» ИТМ № 6/2010) – ред.

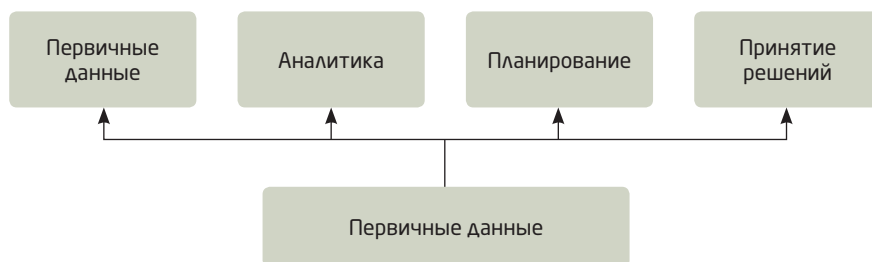


Рис. 2. Общий пул информации

нения решений, а горизонталь – все первичные данные сведем в один большой пул, и будем оттуда выбирать все, что нам надо. См. рис. 2. Даже в нашей модели объем обрабатываемой информации составил 8 уровней по 10 источников – 108. 100 000 000 – сто миллионов первичных документов. Получилось целое море-океан, от горизонта до горизонта, под названием «база данных».

И если на информационной вертикали каждый уровень обработки информации имел дело с небольшим объемом данных, зная специфику этих данных, зная особенности обработки информации своего уровня, то теперь – «компьютер все сделает».

ИС для обработки данных по ДТП

Хмельницкий национальный университет в содружестве с ГАИ много лет успешно ведет работы по созданию и совершенствованию системы обработки данных по ДТП. Не нужно объяснять важность этой задачи, ведь каждое ДТП – это чей-то материальный убыток, многие ДТП – это смерть и горе. Цена правильно принятых решений в этой «войне на дорогах» – спасенные жизни. Ошибка в принятии решений – погибшие, травмированные и в итоге – покалеченные судьбы.

Поэтому в задаче учета ДТП самое важное – не учет карточек и не отчеты перед начальством о том, сколько и кого в этом году задавило, а помощь руководству в верном принятии решений.

В начале разработки было принято решение использовать модель вертикали поддержки принятия реше-

ний, а не «все в кучу, там разберемся». Для нас в задаче с ДТП главное – помочь руководству принимать правильные решения.

Учитывая имеющиеся ресурсы и недостаточные объемы финансирования, была выбрана трехуровневая модель для информационной базы. См. рис. 3. Три уровня на серверах совсем не значит, что нет планирования и принятия решений, просто они делаются уже за пределами нашей системы,



Рис. 3. Трехуровневая модель информационной базы

мы же обрабатываем материал только на первых трех этапах вертикали.

На первом уровне у нас первичный документ со всеми ошибками и опечатками. См. рис. 4. Спецификой первого уровня является то, что используется не реляционная структура данных, а информационная структура, построенная на математической модели многомерного информационного пространства. Это когда каждая единица информации имеет свою уникальную координату в многомерной матрице. В данной задаче мы использовали три координаты. Сходные многомерные структуры применяются очень и очень давно, достаточно взглянуть на Библию, Коран или любой иной древний текст. «Книга Экклезиаста, Глава 1, Стих 10» – 1,10 – координата фразы «Ничто не ново под солнцем». Причем ссылка на 1,10 будет работать, каким бы образом и на каком бы носителе не была представлена эта информация. Вырубите на камне, это будет 1,10. На экране компьютера, в юбилейном издании, в цитатнике – это все равно 1,10. В то же время, если вы сошлетесь на «Библия, издание Сытина от года 1908 от р.х., в мягкой обложке, для простолудиново, по копейке, страница сто третья, сверху пятый абзац» – то узнать, что там написано, сможете только заполучив сие редкое издание, утерянное в веках. Вот это и есть различие ссылки на информацию, от ссылки на носитель инфор-

Номер картки	9001442
Дата заповнення	28.12.2010
Дата останнього редагування	13.01.2011 22:11:37
Тип картки	2
Дата скоєння	28.12.2010
Час скоєння (годин)	11
Час скоєння (хвилин)	11
Вид пригоди	2
Зображення зіткнення	
Підрозділ що оформив ДТП	11008
Підрозділ що обслуговує територію	11007
Картку склав	ПЕТРОВ
ТЛФ	
Резервне (не обов'язкове) поле	
№ ЖРЗПЗ (або книги ДАІ)	111
Дата ЖРЗПЗ (або книги ДАІ)	28.12.2010

Рис. 4. Форма для ввода первичного документа

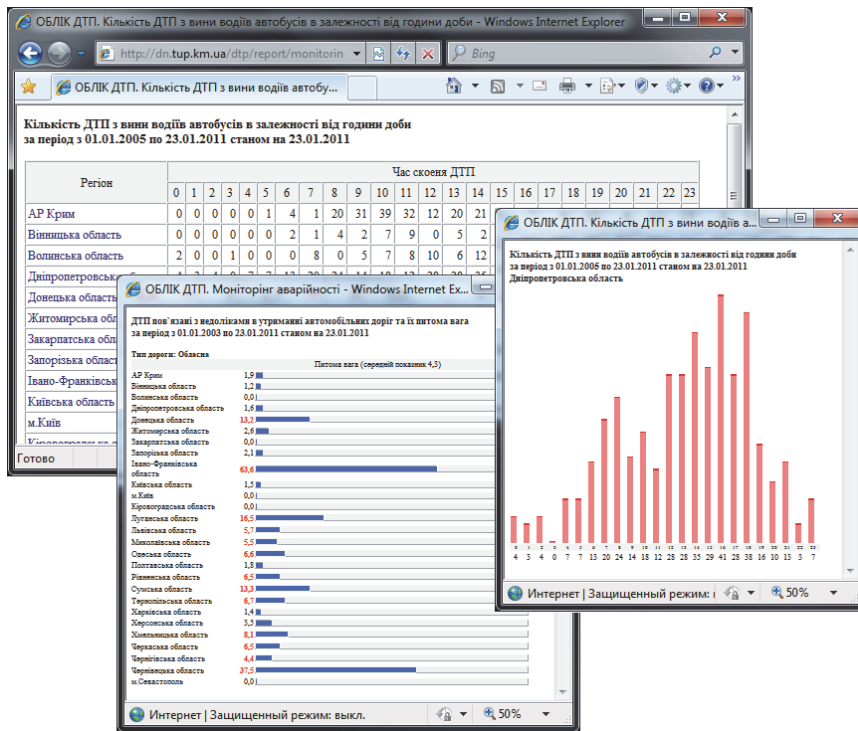


Рис. 5. Примеры статистических отчетов на 2-м уровне

мации. Интернет построен на ссылках на носители, а не на информацию, что во многих случаях существенно усложняет поиск нужной информации.

На уровне первичного документа возможен только ввод и редактирование информации. Круг доступа ограничен. Данные хранятся в оперативном доступе всего два года, перемещаясь потом в архив.

Второй уровень – статистика. См. рис. 5. Формируется статистика раз в сутки, на основе первичных документов. Для удобства пользователей, а работают они по запросам именно на уровне статистики, структура привычная, реляционная – так пользователям проще. Обратите внимание, что пользователи имеют доступ к данным именно на уровне статистики – такой подход гарантирует, что даже очень большой «умник-взломщик» сможет повредить только вторичные данные, которые все равно быстро и просто будут восстановлены по первичным документам.

Статистика отличается от первичного документа не только структурой, но и содержанием. В ней есть некоторые расчетные поля, введенные для удобства пользователя, но отсутствуют некоторые поля, которые ви-

деть обычным статистикам не надо (например, отметки о работе оператора ввода).

На уровне статистики делаются выборки, табличные отчеты, строятся графики, проводится текущий мониторинг ситуации – рутина в рамках ограниченного периода времени без построения трендов и углубленного анализа. Данные накапливаются за пять лет.

Третий уровень возвращается опять к структуре многомерной матрицы, т.к. на этом уровне непосредственно с данными человек не работает. Данные по аналитике формируются раз в месяц, процесс этот для серверов тяжелый и длительный. См. рис. 6.

Именно на аналитическом уровне, т.к. он охватывает большие периоды, можно и вычислить тренды, и сделать прогнозы. Данные аналитики – это «значения по направлениям», а не «Петров Сидорову в первых числах мая», потому цифр в аналитике не много, база не большая, и цифры можно накапливать практически вечно – у них нет срока, после которого данные переходят в архив.

Покажем еще раз на примере, как отличаются данные на трех уровнях.

1. Первичные данные: «Водитель Петров, находясь в состоянии алкогольного опьянения, совершил наезд...»
2. Статистика: «Вид ДТП: 5, нарушение: 1, участник: ПЕТРОВ»
3. Аналитика: «Количество нарушений с кодом 5: 30, что составляет рост 30% и положительную тенденцию на протяжении...»

Первичный уровень описывает ситуацию, как она есть, вплоть до факты и фото. На втором уровне данные сохранены, но формализованы, на тре-

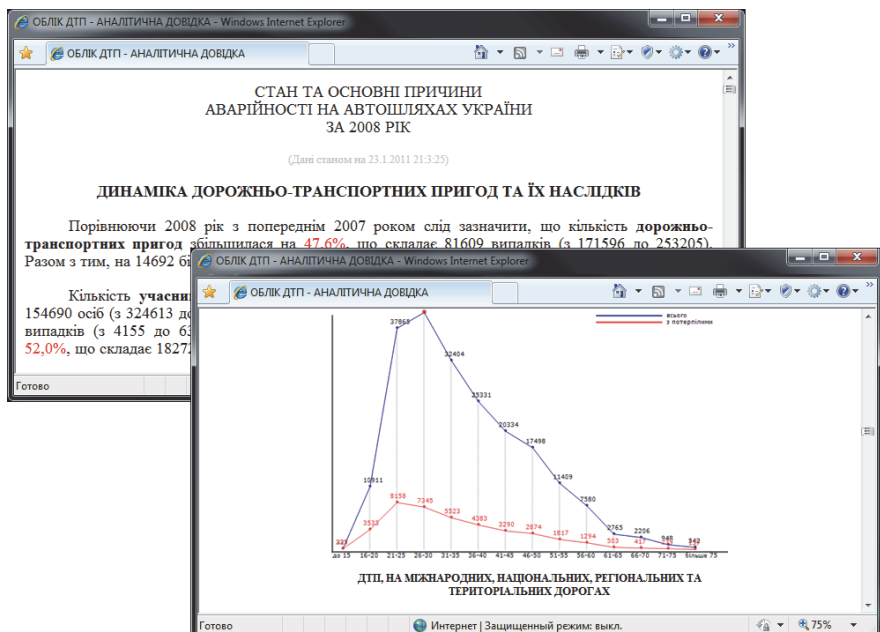


Рис. 6. Примеры аналитических отчетов

тем уровне данные группируются, и никакого ПЕТРОВА уже не найдете, но увидите, какую долю составляют пьяные в ДТП и, самое главное, как эта доля меняется с годами.

Так почему же 2+2 далеко не всегда равно 4

«Девяносто девять источников сообщило, что завтра война, а один агент доложил, что форточка открыта» – какой вывод вы сделаете? Для правильного вывода надо знать, что это за «99 источников» и что имел в виду один агент, упоминая форточку. Формально у нас вероятность войны 99%, а форточку игнорируем. А ведь вполне возможно, что 99 источников совершенно не знают ситуации, тогда как один агент владеет достоверной информацией.

Обработывая информацию, в одних ситуациях необходимо применять математику, а в других – человеческий фактор (на вертикали поддержки принятия решений). Вообще, работа с людьми, агентурной сетью или даже операторами ввода информации – хлопотное занятие, но надежда на «всех заменит компьютер» тщетна. Выбирая свой путь, не торопитесь следовать «моде», как это делают все. Узнайте, как проблемы решались ранее («ничто не ново...»).

Михаил Яновский

Хмельницкий национальный университет