

Визуальные методы работы со знаниями: попытка обзора¹

Аннотация. Статья посвящена обсуждению современного состояния методов структурирования знаний. Вводится классификация наиболее популярных методов. Акцент в обсуждении ставится на визуальных техниках, являющихся мощным когнитивным инструментом. Также рассмотрена методология объектно-структурного анализа как метода стратификации знаний.

Ключевые слова: структурирование знаний, визуальные методы, стратификация знаний.

Введение

Стадия концептуального анализа или структурирования знаний традиционно является (наряду со стадией извлечения) «узким местом» в жизненном цикле разработки интеллектуальных систем (ИС) [10]. Методологии структурирования в основном опираются на современную теорию больших систем или сложных систем [7, 29], где акцент ставится на стадии проектирования.

Среди первых сторонников исследований по теории систем наиболее заметными были Берта-ланфи [21], Раппопорт и Боулдинг [22]. Аналогичные концепции, но связанные не с общесистемными исследованиями, а рассматривающие информационные процессы в системах, таких, как связь и управление, положили начало кибернетике как самостоятельной науке [4, 19]. Этот подход был существенно поддержан работами Шеннона по математическому моделированию понятия информации [18, 23, 28].

Позднее, в 1960-х гг. было сделано несколько попыток сформулировать и развить математические теории систем высокого уровня общности [27]. Существенный вклад в математическую теорию систем и основы структурирования внесли отечественные исследователи Моисеев Н.Н. [14], Глушков В.М. [8], Ивахненко А.Г. [12], Поспелов Д.А. [16], Осипов Г.С.

[15] и другие. Системный анализ тесно переплетается с теорией систем и включает совокупность методов, ориентированных на исследование и моделирование сложных систем — технических, экономических, экологических и т.п. Большой вклад в эту теорию внесли классики объектно-ориентированного анализа [1].

Разработку интеллектуальных систем с уверенностью можно отнести к данному классу задач, поскольку они обладают основными признаками сложности (иерархия понятий, внутриэлементные и межэлементные связи и пр.). Сложность проектирования ИС определяется в основном сложностью знаний предметных областей и управления процессом разработки, а также сложностью обеспечения гибкости конечного программного продукта и описания поведения отдельных подсистем.

1. Классификация методов

В основе любой классификации лежит некоторый признак, и важность выбора правильного принципа структурирования не вызывает сомнений. Ошибки в выделении набора атрибутов-признаков основных объектов предметной области могут свести на нет все преимущества самой информации как таковой. Важны и форма описания, и выделение мета-понятий, категоризация, порядок элементов.

¹ Работа поддержана грантом РФФИ 08-07-00062-а



Рис.1. Классификация методов структурирования знаний

Достаточно условно разделим методы структурирования знаний (МСЗ) на визуальные и символьные (Рис.1).

Символьные методы опираются на математическое, либо текстовое, либо табличное описание знаний, и чаще применяются в хорошо структурированных предметных областях (ПрО). Гораздо хуже дело обстоит со слабо структурированными областями, где имеется большая вариация методов, обусловленная сложной спецификой знаний каждой из областей, оттого предлагающей свои методы и подходы. Часто при этом используются графические или визуальные подходы как менее формальные, наглядные и простые. Использование визуальных методов имеет более древние корни, чем символьное, которое, видимо, и «выросло» исторически из первых графических пиктограмм. Рис. 1 включает далеко не полный список методов, который практически является открытым в части конкретных техник структурирования.

2. Символьные методы

Текстовые методы структурирования знания являются достаточно традиционными, внешне простыми, и часто не вызывают вопросов. В

качестве примера можно взять любой учебник по экономике. Однако сам текст не является знанием, его необходимо «сжать» до уровня чистого смысла, исключив общие места, контекст, заимствования и повторы. Примером, такого сжатого текста является, например, набор правил либо семантическая сеть. В действительности, даже формулирование словесных правил требует неявного, но весьма трудоемкого составления таблиц решений, и в некотором смысле, может быть отнесено к табличным методам.

Применение формул и математического аппарата возможно только в «жестких» хорошо структурированных ПрО, которые не являются «интересными» для интеллектуальных систем, ориентированных на «мягкие» предметные области. Фактически, для баз знаний интеллектуальных систем формулы, как метод концептуализации, используются крайне редко.

Но уже простейшее реферирование текста требует частичного применения специальных техник извлечения знаний, описанных в [6] и требует серьезных аналитических навыков и лингвистических способностей. Мы предлагаем назвать этот прием системно-аналитическим редактированием (САР). В основу САР положен принцип: «Выделить главное и выкинуть

балласт». Метод САР позволяет не только производить компрессию текста, но и вносит в смысл текста структуру и упорядоченность.

Табличные методы требуют двумерной систематизации, т.е. разбиения на группы, подгруппы и иные категории. Количество примеров табличных методов может быть расширено. При этом таблицы могут быть «плоскими», где столбцы и строки внутри себя равноправны, как, например, в реляционной базе данных. А могут быть стратифицированными, подразумевающая неявную или явную иерархию. Понятно, что плоские таблицы составлять легче, хотя любой архитектор базы данных (БД) знает насколько это творческий процесс, требующий серьезной аналитической работы, т.к. выявление атрибутов – суть извлечение и структурирование знаний. Методы разработки БД хорошо описаны [13], в то время как формализация знаний по-прежнему делается методом *ad hoc* (т.е. применительно к случаю).

3. Визуальные методы

Визуализация всегда считалась мощным инструментами познания (*mind tool*), т.е. средством, предназначенным для организации и облегчения процесса познания. Любые модели от плана садового участка до системы дифференциальных уравнений являются инструментами познания, если они помогают нам сформулировать и объяснить природу и структуру явлений. Визуальные модели (например, графы) обладают особой познавательной силой, фактически представляя средства когнитивной графики [11] для структурирования информации. Любой программный графический пакет от PaintBrush до Visio можно использовать как инструмент познания. Однако, сейчас все большую популярность приобретают специализированные методы и инструменты.

На практике, используются более сотни методов визуального структурирования – от традиционных диаграмм, «и-или» графов, сетей Петри, диаграмм бизнес-процессов до модных «стратегических» карт (*roadmaps*), лучевых схем-пауков (*spiders*) и каузальных цепей (*causal chains*). Такое многообразие обусловлено существенными различиями в природе, особенностях и свойствах знаний различных предметных областей. Сегодня любая попытка

перечисления методов визуализации информации с целью структурирования практически обречена на провал. Всегда можно найти новый метод, не вошедший в список. Далее кратко опишем несколько достаточно новых, но уже популярных методов.

Идея интеллект-карт (*mind maps*) заключается в использовании и совмещении функции левого и правого полушарий для достижения целостного и наглядного представления идеи. Фактически, это переход от последовательного (текстового) изложения к сетевому (образному). И-карта – графическое выражение процесса радиантного мышления. И-карты имеют 4 отличительные черты [2]:

- объект внимания/изучения кристаллизован в центральном образе;
- основные темы, связанные с объектом изучения, расходятся от центрального образа в виде ветвей, которые поясняются ключевыми словами или образами;
- вторичные идеи также ветвятся;
- ветви формируют связанную узловую структуру.

Если И-карты показывают связи и древовидную структуру произвольных фрагментов знаний, то концептуальные графы или карты (*concept maps*) позволяют глубже рассмотреть предметную область, и включают отношения между понятиями или концептами. Такие концептуальные графы (К-графы) состоят из узлов и направленных поименованных отношений или связей, соединяющих эти узлы. Связи могут быть различного типа, например, "является", "имеет свойство" и т. п. Концепты и связи имеют универсальный характер для некоторого класса понятий предметной области. Поэтому любая разработка К-графа подразумевает анализ структурных взаимодействий между отдельными понятиями предметной области. Идея К-графа легко поддается формализации в базах знаний ИС в виде семантических сетей [15].

Карты знаний (З-карты) используются в системах управления знаниями предприятий для визуального отображения местоположения источника знания, «владельца» и типа знания. За рамки данной статьи выходит рассмотрение моделей бизнес-процессов [17], онтологий [9] и других методологий структурирования.

Хороший аналитик владеет широкой палитрой методов, всякий раз подбирая метод, наи-

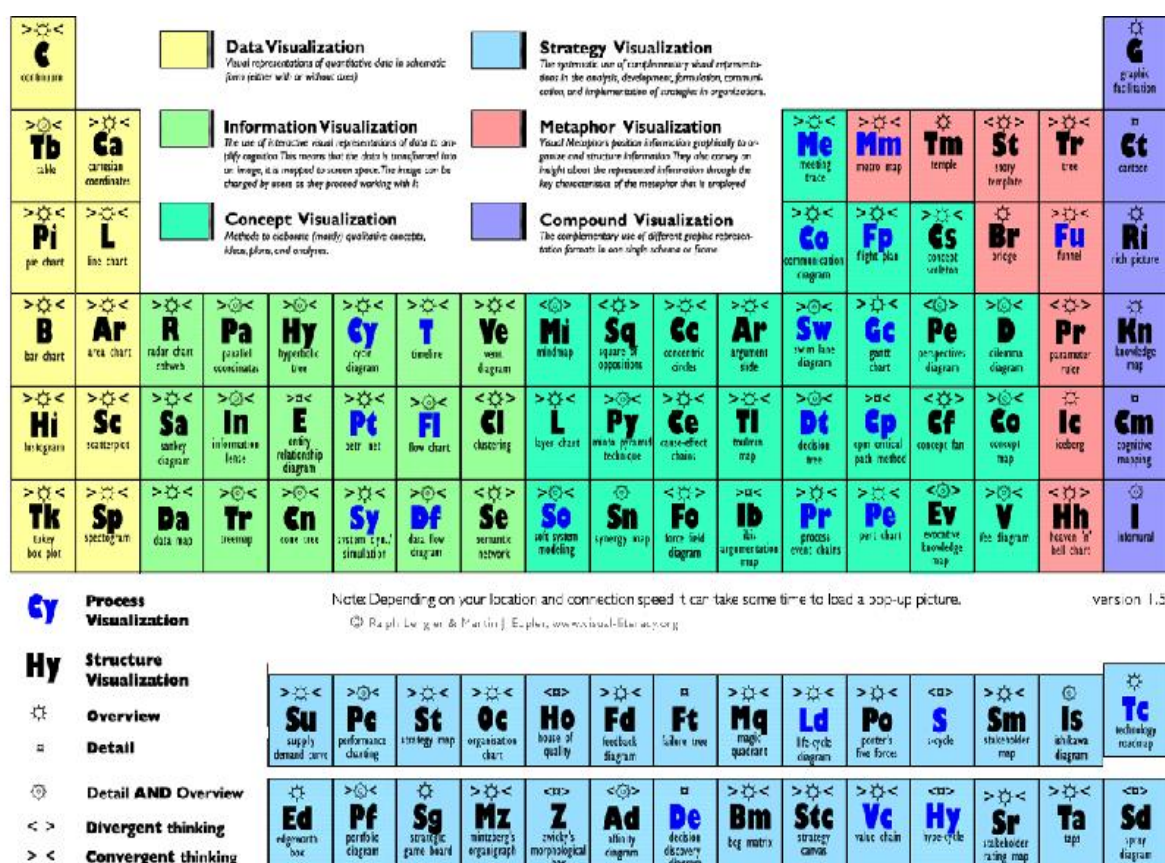


Рис.2. Таблица Ленглера и Эпплера

более адекватный рассматриваемой предметной области. Естественным желанием является стремление разработать хотя бы простейшую классификацию и навести порядок в этом изобилии методов.

На Рис. 2 представлена одна из многочисленных попыток классификации методов визуализации, выполненная Ленглером Р. и Эпплером М. в [26].

Их подход внешне выглядит систематическим, напоминая таблицу Менделеева, однако, по сути он просто описывает методы по нескольким важным признакам.

А) Природа объектов визуализации (данные, информация, концепты, стратегии, метафоры и комплексы) выделены цветом фона ячейки таблицы.

Б) Характер объекта (структура или процесс) показан цветом двухбуквенной метки.

В) Уровень грануляции (обзорный или детальный) отмечен символами ☀ и ■.

Г) Характер мышления (дивергентный и конвергентный) помечен символами < > и > <.

Последний признак требует пояснений, под дихотомией «дивергентное - конвергентное мышление» понимается разница в конкретном (конвергентном) и творческом или креативном (дивергентном) способах мышления [20, 24].

Можно предложить и другие признаки для классификации, например, по типу знания согласно методологии ОСА - объектно-структурного анализа [5]. ОСА подразумевает разбиение ПрО, как правило, на восемь слоев или страт (от лат. *Stratum* - уровень).

Табл. 1. Стратификация знаний предметной области

Номер страты	Тип знания	Характер анализируемой информации
s_1	ЗАЧЕМ-знания	стратегический анализ: цели, назначение и функции системы
s_2	КТО-знания	организационный анализ: человеческие ресурсы, например, коллектив разработчиков системы
s_3	ЧТО-знания	концептуальный анализ: основные концепты, понятийная структура

s_4	КАК-знания	<i>функциональный анализ</i> : гипотезы и модели принятия решения
s_5	ГДЕ-знания	<i>пространственный анализ</i> : окружение, оборудование, коммуникации
s_6	КОГДА-знания	<i>временной анализ</i> : временные параметры и ограничения
s_7	ПОЧЕМУ-знания	<i>каузальный или причинно-следственный анализ</i> : формирование подсистемы объяснений
s_8	СКОЛЬКО-знания	<i>экономический анализ</i> : ресурсы, затраты, прибыль, окупаемость

Таким образом, сначала знания стратифицируются (Табл.1), а затем выбирается модель наиболее подходящая к соответствующему типу знаний (Рис.3).

4. Инструменты структурирования

Однако все большее значение в настоящее время приобретает программный инструмент визуализации. От его выбора часто зависит наглядность и адекватность структуры предметной области. На Рис. 4 представлены некоторые названия инструментов. Хаотичность рисунка подчеркивает разнородность, фрагментарность и несопоставимость элементов из многообразия современных визуальных инструментов.

Для того чтобы упорядочить это множество можно предложить набор признаков для классификации инструментов (Рис. 5).

Рисунок отражает только верхние уровни, описывающие мета-признаки. Продолжение классификации позволит выделять более тонкую специфику задачи, основанную на глубоком понимании особенностей процесса структурирования.

Важными факторами при выборе являются:

1) Функциональность инструмента

Пользователь (на кого рассчитан инструмент)

Профессионал

Новичок

Поддержка коллективной работы (сколько человек может одновременно разрабатывать модель)

Один пользователь

Много пользователей

Расширяемость (возможность внесения изменений)

Масштабируемость

2) Модели (характеристики используемых моделей)

Количество моделей

Тип модели (Рис. 1)



Рис.3. Классификация MC3 по типу знаний согласно OSA



Рис.4. Многообразие инструментов структурирования

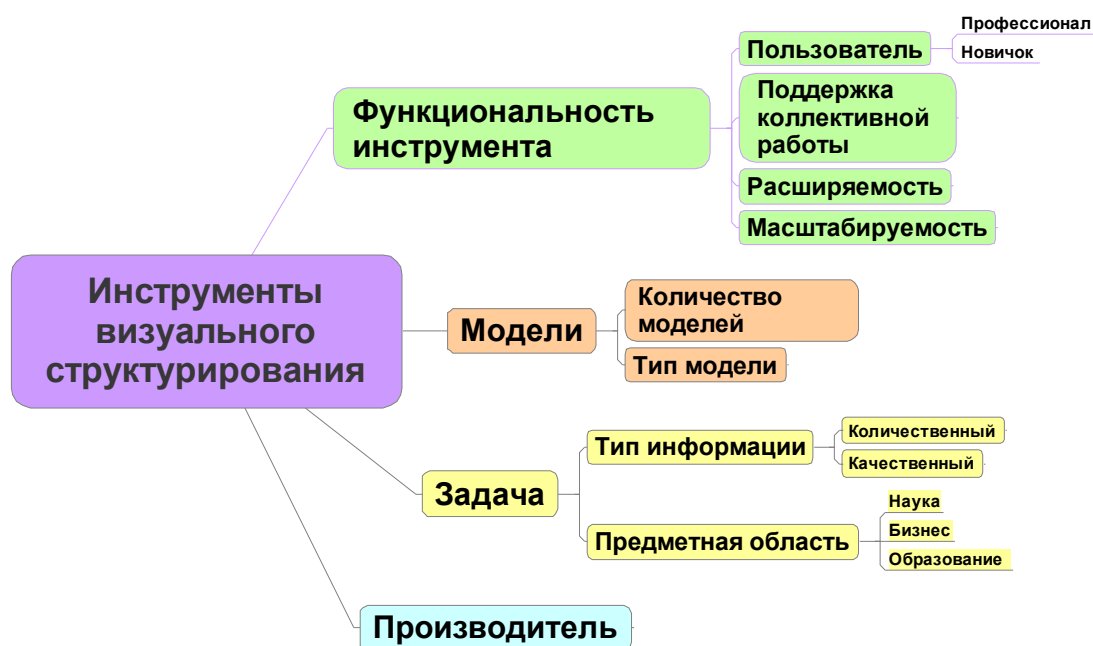


Рис. 5. Основные признаки для классификации инструментов визуального структурирования

3) Задача (характеристики решаемой задачи)

Тип информации

- Количественный
- Качественный

Предметная область

- Наука
- Бизнес
- Образование

4) Производитель (характеристики самого продукта)

Фирма-производитель

- Отечественный
- Иностранный
- Тип продукта

Коммерческий

- Учебно-демонстрационный
- Условно бесплатный

Заключение

Данная статья отражает раннюю стадию процесса осмысления современного состояния моделей и инструментов визуального структурирования информации. Это лишь попытка обзора, безусловно, поверхностная и фрагментарная. Однако, уже сейчас можно констатировать, что не имеет смысла говорить о лучшем инструменте, можно говорить о выборе адекватного и удобного средства согласно задаче, типу информации и квалификации аналитика. Как справедливо заметил ведущий специалист по моделированию бизнес-процессов Г.Н. Калянов: «При выборе лучшим является тот язык моделирования, на котором ты хорошо умеешь работать».

Литература

- Буч Г., Рамбо Дж., Якобсон А. UML. Классика CS, Питер, 2006.
- Бьюзен Б. Супермышление. Белорусский дом печати, Минск, 2002.
- Величковский Б.М. Когнитивная психология. М.: Наука, 1987.
- Винер Н. Кибернетика или управление и связь в животном и машине, М., Сов. радио, 1958.
- Гаврилова Т.А. Объектно-структурная методология концептуального анализа знаний и технология автоматизированного проектирования баз знаний // Труды Междунар. конф. "Знания-Диалог-Решение 95". - Ялта. - 1995. - т.1. - с.1-9.
- Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. Учебник.- СПб, Изд-во «Питер», 2001.
- Гиг Дж. Прикладная общая теория систем, М., Мир, 1981.
- Глушков В. М. Введение в кибернетику, Изд. АН УССР, Киев, 1964.
- Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д. Онтологии и тезаурусы. Казань, КГУ, 2006.
- Джарратано Д., Райли Г. Экспертные системы. Принципы разработки и программирование, Изд-во «Вильямс», 2006.
- Зенкин А.Н. Когнитивная компьютерная графика. – М.: Наука, 1991.
- Ивахненко А. Г. Системы эвристической самоорганизации в технической кибернетике, Изд. Техніка, Киев, 1971.
- Крэнке Д. Теория и практика построения баз данных. 9-е изд. Питер 2004.
- Моисеев Н.Н. Математические задачи системного анализа, М., Наука, 1981.
- Осипов Г.С. Приобретение знаний интеллектуальными системами. М., «Наука», 1997.
- Поспелов Д. А. Ситуационное управление: теория и практика, М., Наука, 1986.
- Тельнов Ю.Ф. Реинжиниринг бизнес-процессов. М.: Финансы и статистика, 2004.
- Шеннон Р., Уивер У. Математическая теория связи. М., ИЛ, 1963.
- Эшби У. Введение в кибернетику, М., ИЛ, 1959.
- Atherton J. Learning and Teaching: Convergent and Divergent Learning 2005.
<http://www.learningandteaching.info/learning/converge.htm>
- Bertalanffy L. An Outline of general Systems Theory // British Journal of the Philosophy of Science, v.1, 1950. – pp.134-164.
- Boulding K. General Systems Theory // Management Science, n.2, 1956. – pp.197-208.
- Feinstein A. Foundation of Information Theory, McGraw-hill, N.Y., 1958.
- Hudson L., Contrary Imaginations; a psychological study of the English Schoolboy Harmondsworth: Penguin, 1967.
- Lengler R., Eppler M. Towards A Periodic Table of Visualization Methods for Management. IASTED Proceedings of the Conference on Graphics and Visualization in Engineering (GVE 2007), Clearwater, Florida, USA, 2007
- Lengler R., Eppler M. (2007) http://www.visual-literacy.org/periodic_table/periodic_table.html
- Mesarovic M. Views on General Systems Theory, Wiley, N.Y., 1964.
- Watanabe S. Knowing and Guessing, Wiley, N.Y., 1969.
- Wasson C. System Analysis, Design, and Development: Concepts, Principles, and Practices (Wiley Series in Systems Engineering and Management), Wiley-Interscience, 2005.

Гаврилова Татьяна Альбертовна, заведующая кафедрой информационных технологий в менеджменте Высшей школы менеджмента Санкт-Петербургского Государственного Университета. Окончила электротехнический факультет Ленинградского института инженеров железнодорожного транспорта в 1975 году. Доктор технических наук, профессор. Имеет более 100 публикаций. Область научных интересов: онтологический инжиниринг, человеко-машинное взаимодействие.

Гулякина Наталья Анатольевна, заведующая научно-исследовательской лабораторией новых информационных технологий Белорусского государственного университета информатики и радиоэлектроники. Окончила Белорусский государственный университет, факультет прикладной математики в 1974 году. Кандидат физико-математических наук, доцент. Имеет более 70 печатных работ. Область научных интересов: инженерия знаний, интеллектуальные обучающие системы и виртуальные организации.