

УДК 004.822

Г.Б. Бронфельд

**НЕКОТОРЫЕ ВОЗМОЖНОСТИ ФОРМАЛЬНОГО ПРЕДСТАВЛЕНИЯ
СТРУКТУРЫ ЗНАНИЙ**

ООО «НИПИ «Сириус-2», Нижний Новгород

Рассматривается популярная 4-уровневая структура модельного представления знаний. В ее состав входят данные (понятия), информация, собственно знания и мудрость. Новым является представление знаний и мудрости в виде путей на графе информации. Это открывает новые возможности аналитического изучения знаний и информации, реализации новых подходов в создании интеллектуальных систем и делает более четкими и наглядными взаимосвязи между информацией и знанием.

Ключевые слова: знание, модель знаний, информация, текст, моделирование, структура знаний, семантическая сеть, база знаний, интеграция знаний

Введение

Все большее внимание в последние десятилетия уделяется работе со знаниями.

Важным моментом в этой проблеме является анализ структуры знаний, поскольку понимание структуры знаний подводит к определенным методам, необходимым для моделирования знаний и выводам на их основе.

Проблематику знаний, в т.ч. во взаимосвязи с информацией, изучают философы, психологи, лингвисты, физики, математики, в рамках ИТ-технологий и направления «искусственный интеллект» (ИИ). Будем вести рассмотрение, в основном, с точки зрения направления ИИ (далее просто ИИ).

Нас будет интересовать научное знание, личное знание и реальная окружающая природа.

Научное знание претендует на адекватное отражение действительности и выступает от имени истины [1].

Личное знание – это совокупность знаний конкретного человека, отражающих его опыт и представления об окружающей действительности, не имеющих, как правило, теоретико-систематическую формы и помогающих ориентироваться в окружающем мире [1].

Действительность – это окружающий нас мир с совокупностью всех существующих объектов, их взаимосвязей и природных явлений [1].

Действительность – объективна и существует независимо от человека. Но человек в любом случае может описывать окружающий мир лишь с помощью языковых средств, фиксируемых, в частности, в виде текстов.

Текст представляет собой последовательность языковых единиц, объединенных смысловой связью [2]. Типичным представителем языковых единиц являются слова, являющиеся некими понятиями, отражающие определенные явления действительности.

Личное знание человек может выразить или опять-таки языковыми единицами, или действиями, которые описываются также с помощью текстов.

В результате получается, что в основе имеем понятия или данные (сведения), отражающие уровень неких параметров действительности (известных или неизвестных нам) и выражаемые в текстовом виде.

Взаимосвязь информации и знаний исследуют по многим направлениям одновременно, и поддерживающие тот или иной подход специалисты нередко разделились на группы и даже друг друга не анализируют и не цитируют.

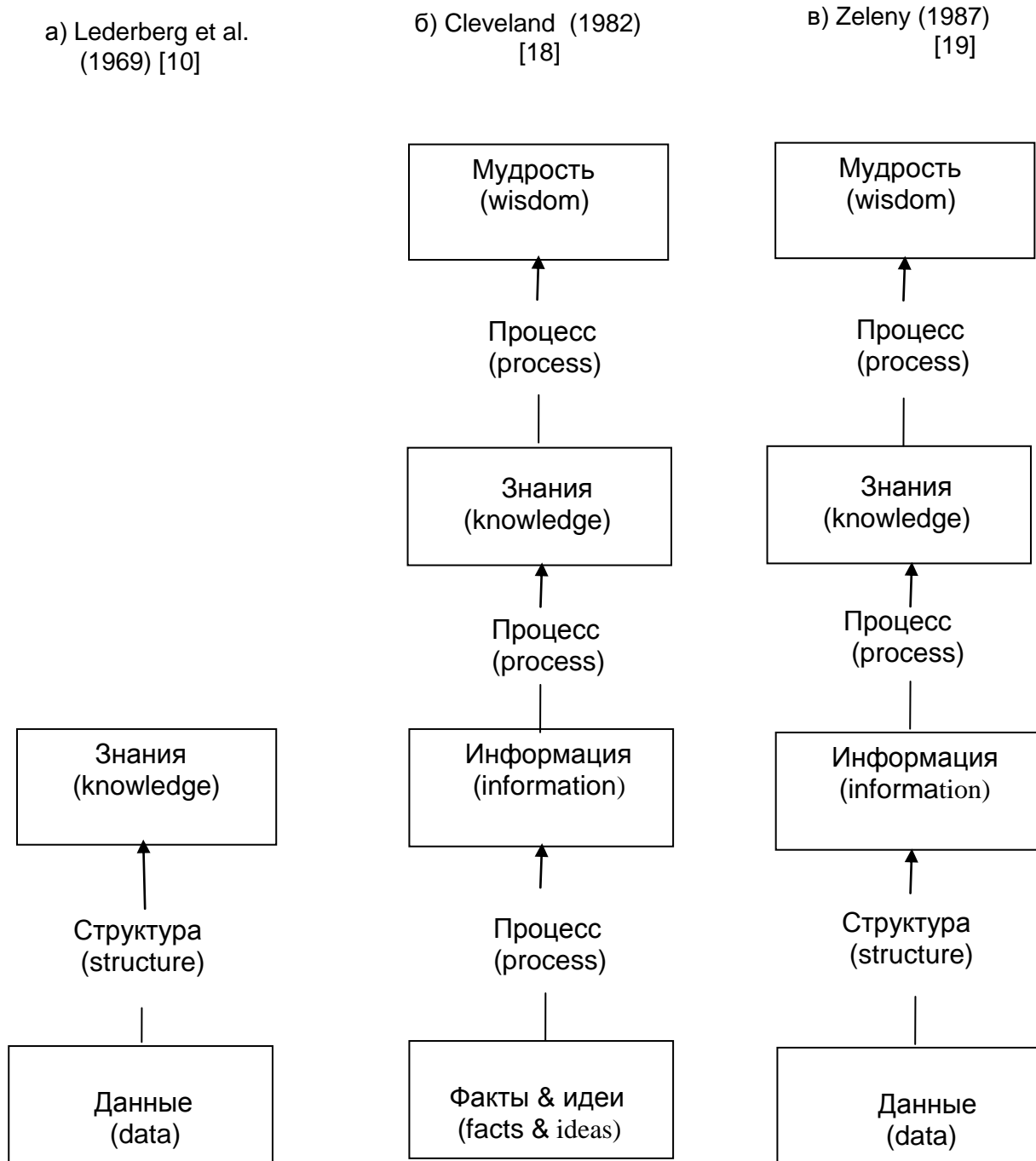


Рис. 1. 2- и 4-уровневая структура знаний

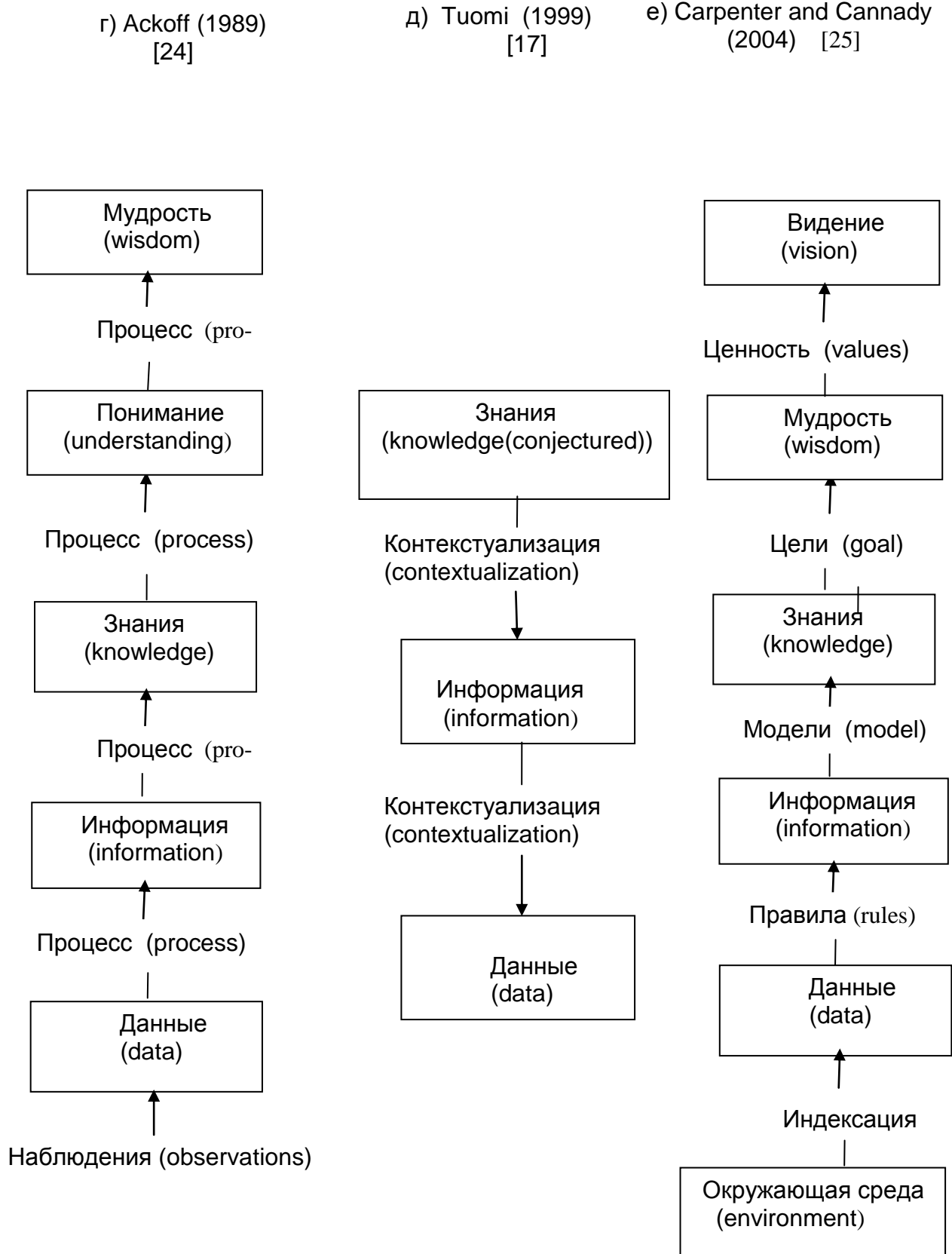


Рис. 2. 3-, 5- и 6-уровневая структура знаний

Одни придерживаются атрибутивной точки зрения [3–6], считая, «что информация предполагается неотъемлемым свойством (атрибутом) материи и поэтому она может прояв-

лять себя во всех объектах, процессах и явлениях как живой, так и неживой природы» [3]. Назовем это определение информацией 1.

Другие рассматривают информацию и знания с точки зрения ее принадлежности высокоорганизованным живым объектам [6–8], назовем информацией 2. Есть активно применяемый для отдельных направлений шенноновский подход на основе статистической оценки количества информации [8], - информация 3. Однако он сам полезен именно для живых существ. Ряд авторов периодически ведет просто полуфилософское вербальное обсуждение этой проблематики на основе интеграционного подхода без серьезного продвижения вперед [9].

Для описания знаний в ИИ давно сформировалась представление о двухуровневой структуре знаний (рис. 1, а), пока наиболее часто используемое в России [10–15]. В ее основе находятся данные (понятия), т.е. отдельные факты, характеризующие объекты, процессы и явления предметной области, а также их свойства. Тогда знания - это хорошо структурированные данные, или данные о данных, или метаданные.

Есть традиционные представления знаний, использующих эту основу, в частности продукционные модели, семантические сети и фреймы. Однако постепенно стал развиваться более широкий спектр структур представлений знаний и информации [16] (рис. 1 и рис. 2). Появилось 3-уровневое представление [17], 4-уровневое [18–23], 5-уровневое [24] и даже 6-уровневое [25]. Эти структуры представлены на рис. 1. В ряде этих структур также используется информация как составной элемент знаний, назовем информация 4. Особое место занимает юридическое определение термина «информация» [26–27], назовем его информация 5.

Для дальнейших теоретических исследований и практических работ необходимо все-таки уточнить взаимосвязи различных трактовок понятия «информация» между собой, с понятием «знание» и все-таки формально уточнить, что же такое данные, информация, знания и мудрость в новых структурных представлениях.

Четырехуровневое представление знаний

Рассмотрим более подробно четырехуровневое представление знаний. Оно и есть относительно новое, менее знакомое, и, на взгляд автора, более интересное представление о системологии знаний во взаимосвязи с информацией, вполне приемлемое для описания научного знания, личного знания и действительности. Затем с позиции новых результатов проанализируем упомянутые структуры снова.

Авторы М.Желена [19], Х.Кливленд [18] и близкий к ним по взглядам Р. Акофф [24]) считают, что знания состоят (рис. 1, в):

- из данных (понятий);
- информации;
- собственно знаний (в узком смысле);
- мудрости.

Данные (понятия) – самые простые элементы знания, их цель неясна и неоднозначна, им свойственна большая степень свободы.

Информации свойственна более высокая конкретность цели, включающей объединение сведений или данных. Как только данные преобразованы в информацию, вернуть их в исходное состояние трудно (так же, как невозможно реконструировать отдельные наблюдения из их средней величины).

Собственно знания охватывают реальный процесс обработки входных данных, включая координацию действия для достижения результатов, целей или получения продуктов.

Тогда мудрость соотносится с объяснимостью: если человек знает почему – а не просто, что и как – тогда он не просто хорошо информирован или осведомлен, а обладает знанием и компетентностью.

Такое представление точно описывает взгляд, в частности, М. Желены [19–21], на знание в следующем виде: «знание – это целенаправленное координированное действие». Его

единственное доказательство или способ демонстрации заключается в достижении цели. Качество знания может быть оценено по качеству достижения (его продукта) или качеству координирования (его процесса), особенно если вмешиваются неконтролируемые факторы.

Хорошо известно уже упомянутое ранее представление моделей знаний в виде двухуровневой структуры (рис. 1, а) с помощью семантических сетей, которые можно описать в виде ориентированного графа $G(V, E)$, где V – множество вершин; E – множество ребер (отношений) между ними. Вершинам соответствуют определенные дефиниции данных (понятий) в виде множества X , дефиниции видов отношений (связей) в виде множества T , связанные отображениями $D: X \rightarrow V$ и $F: T \rightarrow E$.

Формальное представление четырехуровневой структуры знаний

Разовьем это формализованное представление двухуровневой структуры знаний (рис. 1, а) до четырехуровневой структуры (рис. 1, в).

Таким образом:

- данные(понятия) – это будут соответствовать вершины V' графа $G(V', E')$ (рис. 3);
- информация – это граф $G(V', E')$, как минимум, с одной ребром (рис. 4);
- знания – это путь в графе, ведущий к некоему результату (вершине) (рис. 5);
- мудрость – это разъяснение (интерпретация), почему выбран именно данный путь из бесконечного множества (рис. 6).

Рассмотрим подробнее.

К данным (понятия), в частности, относятся вершины $v_1, v_2, \dots, v_i, \dots, v_k \in V$

(рис. 3), которые задаются непосредственно. Кратко рассмотрим это на примере авторского изобретения комбинированного устройства «письменный стол» (ПЭВМ) [28], который в настоящее время массово производится в мире под названием *powerdesk*. С виду столем («*powerdesk*») почти стандартный письменный стол с тумбочкой со свободной поверхностью стола. В конструктивы письменного стола встроены составляющие ПЭВМ. Это создает большие удобства для работы человека и позволяет со временем быстро проводить модернизацию (замену отдельных элементов ПЭВМ) не меняя удобное устройство для пользователя в целом. К числу данных (понятий) для данного случая, могут относиться, - например, процессор, клавиатура, дисплей, древесно-стружечная плита, роликовые опоры и т.д. Однако при решении конкретной задачи может потребоваться неограниченное дополнительное множество данных (понятий) V_δ , причем $V \cup V_\delta = V'$ т.е. существует множества исходное фиксированное X и дополнительное X_δ , соответственно, и связи фиксированные исходно T и дополнительное T_δ , и $D_\delta: X_\delta \rightarrow V_\delta$. Соответственно $D \cup D_\delta = D'$:

Как научное, так и личное знание состоит из дополнительных данных и понятий X_δ (с их связями T_δ), имеющих отношение к действительности, и необходимых для решения возникающих проблем.

Информация (рис. 4) включает в себя ребра $v_1v_4, v_2v_4, \dots, v_3v_k \in E$, по примеру, это элементы чертежа стола, описание устройств ЭВМ и т.д. Но ведь все реально шире. Например, размер конструкции определялся эргономическими параметрами человека, и, в частности, возможностями, поднять столем и перевозить. Значит появляется информация, о человеке, автомобилях, способах перевозки, методах упаковки, правилах внешнеэкономической деятельности и торговли (приходилось и в другие страны продавать) и т.д. Т.е. рядом с базовым набором данных V неявно появляется дополнительный набор данных (понятий) V' , которые как-то связаны с исходными данными (понятиями),

$$v_1v_{k+1}, v_{k+1}v_{k+2}, v_2v_{k+2}, v_{k+2}v_{k+4}, v_{k+2}v_k, v_{k+4}v_3, v_{k+2}v_k, v_kv_{k+3}, v_3v_{k+5}, v_{k+5}v_n \in E_\delta.$$

Эти связи могут относиться, как к эксплицитным, так и имплицитным знаниям (касающимся личного знания) человека [14, 20].

Тогда $E \cup E_0 = E'$, что соответствует $F_0 : T_0 \rightarrow E_0$ и $T \cup T_0 = T'$.

Собственно знания (рис. 5) в данном примере являются технологическими картами по изготовлению составных элементов стола, их сборке, отладке и проверке работоспособности и качества. Это когда решение известно, но если решение неизвестно? В столе как в новое изобретение были заложены некие новые понятия, связи и решения, которые были неизвестны до того (но также формировались на основе известного до этого периода – это человеческая особенность). Кроме того, если решение (знание) предварительно неизвестно в процессе разработки проблемы, могут уточняться и сама проблема и способы ее решения [29,30], причем этот процесс итеративный.

На рис. 5 представлено это набор конкретных путей от одного множества начальных вершин $V^0 \subset V'$ к множеству конечных вершин $V^R \subset V'$, в частности, путь

$$s_i = v_1 v_4 v_2 v_{k+2} v_{k+4} v_3 v_k \in S,$$

где $v_1 \in V^0$, а $v_k \in V^R$

Мудрость (рис. 6) в данном примере заключается в объяснении, зачем нужен стол, в объяснении его достоинств в эксплуатации по сравнению с другими аналогичными видами компьютерной мебели. Мудрость – это множество интерпретаций M_{si} , поясняющих полезность i -го пути s_i , в частности,

$$m_{si} = v_{k+1} v_{k+2} v_k v_3 v_{k+5} \in M_{si}.$$

Следует отметить, в объяснении m_{si} могут использоваться данные (понятия), в большинстве не участвующие в исходном описании со множеством V , но входящие во множество V_0

Тут надо сделать замечание, что граф $G(V', E')$ может в общем виде иметь вид гиперграфа с гораздо более сложными видами связей, а сами пути s_i и m_{si} могут иметь характер маршрутов с наличием замкнутых циклов с повторением вершин и ребер. В статье рассматривается более простой вариант. Более сложное представление сути обсуждаемого вопроса не меняет.

Полное формальное описание четырехуровневой структуры знаний имеет вид семерки

$$ZN = \langle X'^0, X'^R, D', F', G(V', E'), S, M_{sk} \rangle, \quad (1)$$

где X'^0 – подмножество начальных данных (понятий);

$X'^0 \subseteq X^0 \subset X \subseteq X'$, относительно которых существуют знания;

X'^R – подмножество конечных данных (понятий) $X'^R \subseteq X^R \subset X \subseteq X'$, относительно которых существуют знания;

D' – определение исходных данных (понятий);

F' – определение связей

$G(V', E')$ – информация;

S – множество путей на графе, которые обеспечивают полезный переход из начального подмножества X'^0 в конечное подмножество X'^R ;

M_{sk} – множество интерпретаций, поясняющих полезность k -го пути s_k .

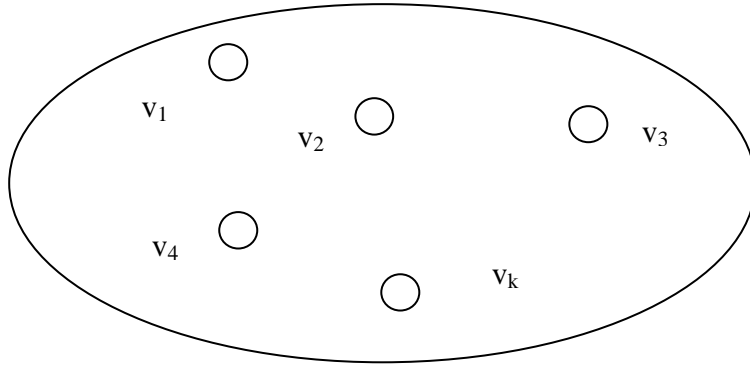


Рис. 3. Данные

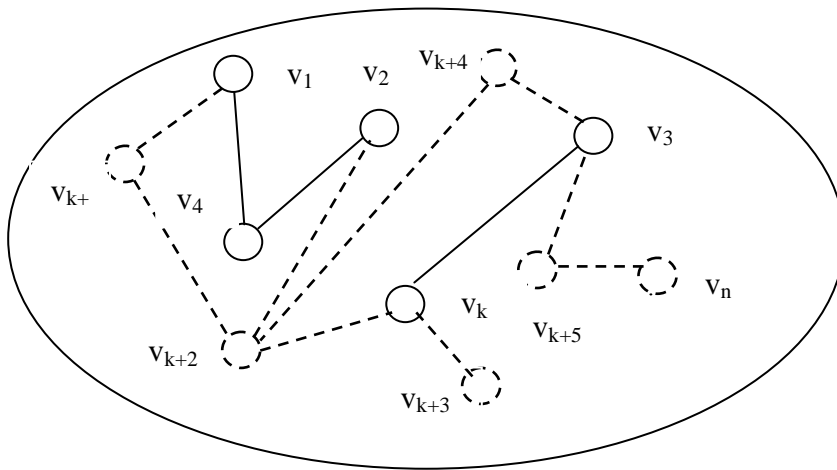


Рис. 4. Информация

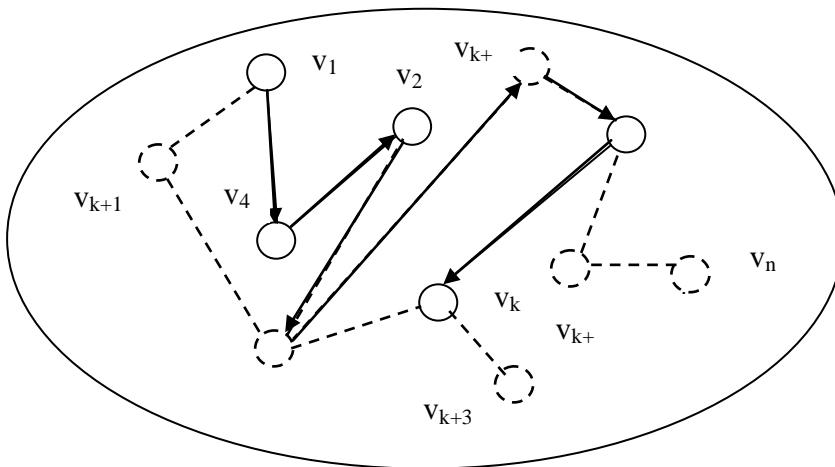


Рис. 5. Знания

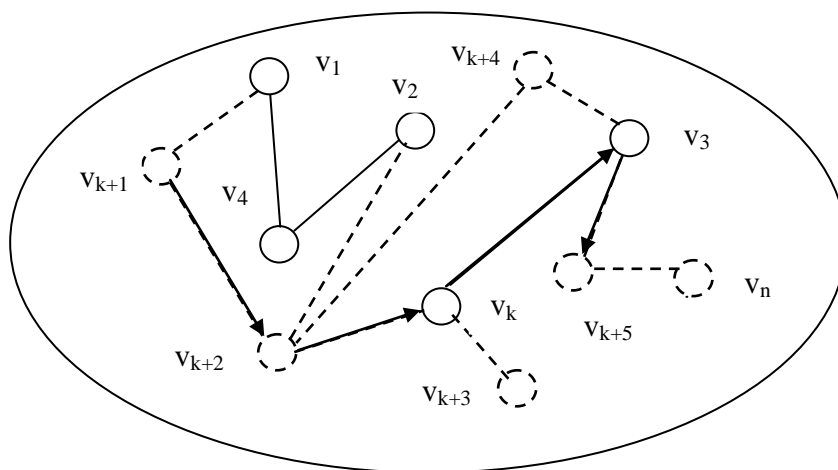


Рис. 6. Мудрость

Информация $G(V', E')$ в (1) у нас соответствует упоминаемому понятию «информация 4». Обратим, внимание, что при определенных условиях выполняется отображение

$$ZN \rightarrow G(V', E'),$$

причем заведомо $G(V, E) \subseteq G(V', E')$. Но тогда $G(V', E')$ можно считать понятием «информация 5», т.е., по сути, знания входят в «информацию 5». Но тогда получается «информация 4» и «информация 5» – одно и то же понятие. Но с определенной разницей, что термин «информация 4» можно трактовать только как граф $G(V', E')$, из которого мы еще не знаем, что является знанием, что мудростью. Понятие «информация 5» включает в себя совокупность: отдельно и описание данных, и $G(V', E')$ и описание всего множества путей S , определяющих собственно знание, и множество путей M_{sk} , определяющих мудрость.

Новая технология [31, 32] по прямому наложению знаний позволяет, в принципе, постепенно создавать реально базы знаний для интеллектуальных систем в рамках ИИ, которые будут уже действительно обладать свойствами графа $G(V', E')$ и включать уже объединенные специальным образом: и просто информацию уровня «информация 4», и знания S , и мудрость M_{sk} .

Это означает: несмотря на то, что исходно информация входит как структурная составляющая знания, в свою очередь, знание можно рассматривать как составляющий элемент информации. Такая трактовка удовлетворяет и понятию «информация 2», но лишь частично связана с понятием «информация 1». А термин «информация 3», связанный со статистической оценкой количества информации, является вспомогательным понятием и имеет возможности совмещения с понятиями «информация 3», «информация 4», «информация 5». Т.е. остаются два разных ключевых подхода, связанных с разными интерпретациями: «информация 1» и «информация 5».

Безусловно, 4-уровневая структура (рис. 1, б, в) имеет больше конструктивных возможностей для развития, чем 2-уровневая, ничего, по сути, не запутывая, а наоборот разъясняя. 5-уровневая структура Акоффа (рис. 2, з) практически ничем не отличается от 4-уровневой и сводится к ней. Взгляд на проблему [17] с 3-уровневой структурой (рис. 3, д) имеет место как методологический взгляд на проблему, который не противоречит 4-уровневой структуре знаний. 6-уровневое представление знаний (рис. 2, е) тоже можно счи-

тать просто методологическим представлением, по сути, ничем не отличающимся от 4-уровневой структуры (рис .1, б, в).

Выводы

Подобная формальная трактовка знаний и мудрости в 4-уровневой структуре знаний пригодна, как для формального описания научного знания, отражающего окружающую действительность, так и для описания личного знания и знаний, содержащихся в текстах на естественном языке. Появляются новые возможности для аналитического углубления в особенности знаний и информации, невозможные до сих пор. Рассмотренная трактовка структуры удобна для разработки новых подходов к созданию интеллектуальных систем в рамках ИИ.

Библиографический список

1. **Кохановский, В.П.** Основы философии науки / В.П. Кохановский [и др.]. – Ростов н/Д: Феникс, 2004. – 608 с.
2. Лингвистический энциклопедический словарь / под В.Н. Ярцева. – М.: Сов. энциклопедия, 1990. – 685 с.
3. **Колин, К.К.** Сущность информации и философские основы информатики // Информационные технологии. 2005. №5. С. 63-70.
4. **Кадомцев, Б.Б.** Динамика и информация / Б.Б. Кадомцев. – М.: Редакция журнала «Успехи физических наук», 1999. – 400 с.
5. **Гуревич, И.М.** Оценка информационных характеристик Вселенной // Информационные технологии. 2008. №12. Приложение. С. 1–32.
6. **Фридланд, А.Я.** О сущности информации: два подхода // Информационные технологии, 2005. №5. С. 75-85.
7. **Винер, Н.** Кибернетика или управление и связь в животном и машине / Н. Винер. – М.: Советское радио, 1968. – 328 с.
8. **Шеннон, К.** Работы по теории информации и кибернетике / К. Шеннон. – М.: Иностранная литература, 1963. – 830 с.
9. **Bates, M.J.** Information and knowledge: an evolutionary framework for information science// Information Research, Vol. 10 No. 4, July 2005, paper 239.
10. Lederberg J., Sutherland G.L., Buchanan B.C., Feigenbaum E.A., Robertson A.V., Duffield A.M. and Djerassi C. Applications of artificial intelligence for chemical inference. 1. The number of possible organic compounds. Acyclic structures containing C,H, O and N//Journal of American Chemical Society, May 1969.
11. Fifth generation computer systems: proceedings of the international Conference on Fifth Generation Computer Systems. - Tokyo, Japan, October 19-22, 1981. – 287 p.
12. **Попов, Э.В.** Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ / Э.В. Попов. – М.: Наука, 1987. – 288 с.
13. **Кандрашина, Е.Ю.** Представление знаний времени и пространстве в интеллектуальных системах / Е.Ю. Кандрашина, Л.В. Литвинцева, Д.А. Поспелов. – М.: Наука, 1989. – 328 с.
14. **Гаврилова, Т.А.** Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб.: Питер, 2001. – 384 с.
15. **Вагин, В.Н.** Знание в интеллектуальных системах // Новости искусственного интеллекта, 2002. №6(54). С. 8-18.
16. Carpenter S. A. A Primer: Enterprise Wisdom Management and the Flow of Understanding, 2008 <http://www.cognitivecybernetics.com/PrimerFoU.html>
17. Tuomi, I. Data is More than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. Journal of Management Information Systems, 16(3) (Winter 1999-2000), 103-107
18. Cleveland H. Information as a Resource// The Futurist, December 1982 p. 34–39.
19. Zeleny M. Management Support Systems: Towards Integrated Knowledge Management// Human Systems Management 7 (1), 1987, p. 59–70.

20. Информационные технологии в бизнесе / под ред. М. Желены. – СПб: Питер, 2002. – 1120 с.
21. Zeleny Milan Human Systems Management: Integrating Knowledge// Management and Systems. World Scientific, 2005, p. 15–16.
22. Bellinger, G. Knowledge Management—Emerging Perspectives. Circa 1997. Last retrieved February 6, 2005: <http://www.systems-thinking.org/kmgmt/kmgmt.htm>
23. **Бронфельд, Г.Б.** Об одном подходе к построению интеллектуальных электронных книг и о молинге // Интеллектуальные системы: труды Восьмого международного симпозиума. – М.: РУСАКИ, 2008. С. 230-234.
24. Ackoff R. From Data to Wisdom//Journal of Applied Systems Analysis,1989, №16, p.3–9.
25. Carpenter, S. A. and Cannady, J. Tool for Sharing and Assessing Models of Fusion-Based Space Transportation Systems. Proceedings of the 40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, (July 11-14, 2004), Fort Lauderdale, Florida.
26. Информационные ресурсы для принятия решений / А.П. Вереvченко [и др.]. – М.: Академический проект, 2002. – 560 с.
27. Об информации, информационных технологиях и о защите информации. Федеральный закон от 27.07.2006 № 149-ФЗ <http://base.garant.ru/12148555/>
28. **Бронфельд, Г.Б.** Комбинированное устройство: письменный стол-ПЭВМ / Г.Б. Бронфельд, А.И. Соболев // Микропроцессорные средства и системы. 1988. № 4. С. 95–96.
29. **Альтшуллер, Г.С.** Найти идею. Введение в теорию решения изобретательских задач / Г.С. Альтшуллер. – Новосибирск: Наука, 1986. – 209 с.
30. **Орлов, М.А.** Основы классической ТРИЗ. Практическое руководство для изобретательского мышления / М.А. Орлов. – М.: Солон-Пресс, 2005. – 416 с.
31. Бронфельд Г.Б. Прямое наложение знаний и некоторые его особенности // Технические и программные средства систем управления, контроля и измерения (УКИ-10): труды российской конференции. – М.: ИПУ им. В.А.Трапезникова РАН, 2010. С. 201–211.
32. Пат. на изобретение, № 2440610 Российская Федерация, МПК G06 N 5/00. Система для работы с интеллектуальной электронной книгой – элингой / Бронфельд Г.Б.; заяв. 21.06.2010; опубл. 20.01.12, Бюл. № 2.

*Дата поступления
в редакцию 05.10.2012*

G. Bronfeld

OPPORTUNITIES OF FORMAL REPRESENTATION FOR KNOWLEDGE STRUCTURE

Purpose: Exploration of correlation of knowledge and information on the basis of new structural and formal representations.

Design/methodology/approach: The paper investigates a popular 4-layer structure of knowledge representation. Its structure contains data (concepts), information, actual knowledge and wisdom. Other structural knowledge representations -- 2-layer, 3-layer, 5-layer, and 6-layer are also considered.

Findings: The paper contains new representation of knowledge and wisdom as ways on the information's graph. It is considered by example of a concrete subject situation. A formal detailed model uses mathematical description and its graphic representation.

Conclusion: This represents new opportunities for analytical study of knowledge and information and realizations of new approaches in creation of intellectual systems. As a result there is more precise and evident correlation between information and knowledge and their false treatments are found out.