



МОДЕЛИРОВАНИЕ СТРУКТУРЫ ПРЕДСТАВЛЕННЫХ В УЧЕБНЫХ ТЕКСТАХ ЗНАНИЙ

М.В. Воронов

Московский государственный психолого-педагогический университет, Россия.

Аннотация. Информатизация образования представляет собой составную часть многогранного процесса развития цивилизации на современном этапе. Одним из актуальных направлений этого процесса является разработка методов формализации знаний, представленных в тексте в целом и в учебной литературе в частности. В статье рассматривается задача трансформации учебного текста, представленного на естественном языке, в цифровой текст. Совокупность представленных в учебных текстах знаний трактуется как система. При этом отмечается, что существенный объем сведений находится в структуре содержащихся в учебных текстах знаний, что с успехом может использоваться при создании средств, обеспечивающих повышение эффективности образовательных процессов. Предлагается метод построения математической модели структуры знаний изложенного на естественном языке учебного текста. Модель структуры знаний позволит обеспечить разработку программно-технических средств, поддерживающих процедуры анализа учебных текстов и получения значений количественных оценок их качества. Обсуждаются возможности применения таких моделей для создания средств активной поддержки самоподготовки обучаемых.

Ключевые слова. Образование, информатизация, процесс, знания, учебный текст, структура, модель.

MODELING OF THE STRUCTURE OF KNOWLEDGE PRESENTED IN EDUCATIONAL TEXTS

M.V. Voronov

Moscow State Psychological and Pedagogical University, Russia.

Annotation. Informatization of education is an integral part of the multifaceted process of development of civilization at the present stage. One of the actual directions of this process is

the development of methods of formalization of knowledge presented in the text in general and in the educational literature in particular. The article deals with the task of transforming the educational text presented in natural language into a digital text. The totality of knowledge presented in educational texts is interpreted as a system. At the same time, it is noted that a significant amount of information is in the structure of the knowledge contained in educational texts, which can be successfully used in the creation of tools to improve the efficiency of educational processes. A method for constructing a mathematical model of the knowledge structure of a natural language teaching text is proposed. The knowledge structure model will allow for the development of software and hardware tools that support the procedures for analyzing educational texts and obtaining the values of quantitative assessments of their quality. The possibilities of using such models to create means of active support for self-training of trainees are discussed.

Keywords. Education, informatization, process, knowledge, educational text, structure, model.

Постановка задачи

Цифровизация образования представляет собой составную часть многогранного процесса развития цивилизации на современном этапе. Современная образовательная среда во все большей мере использует средства, которые на основе информационно-коммуникационных технологий поддерживают осуществление все большего количества образовательных функций. Тем самым появляются возможности более эффективно решать традиционные задачи и задачи, которые ранее могли быть решены только при огромных затратах живого труда. Среди актуальных в области информатизации сферы образования выделяются задачи содержательной обработки текстов.

Представленный на естественном языке текст и в настоящее время остается важнейшим средством фиксации, хранения и передачи информации [1]. Именно текст является основой подавляющего большинства учебных ресурсов. Стремление же использовать текст в компьютерных средствах поддержки образовательных процессов сталкивается с необходимостью преобразования учебного текста, представленного, как правило, на естественном языке в так называемый цифровой текст (текст на «машинном языке»), т.е. текст, продуцированный с помощью какого-либо электронного носителя информации.

Современные информационные технологии позволяют эффективно трансформировать исходные тексты в тексты цифровые, чтобы в последующем с помощью компьютерных средств оперировать ими. Помимо хранения, переноса и представления текста в удобной для обработки и прямого использования форме созданы

средства, обеспечивающие для текстов, представленных на естественном языке, возможность производить информационный поиск, извлекать именованные сущности и факты, создавать вопросно-ответные системы, анализировать тексты и др. [2]. Вместе с тем исследования в области цифровизации текста активно продолжаются и в значительной мере направлены на то, чтобы при различных трансформациях цифровой текст в наибольшей степени сохранял содержание исходного текста.

В этом направлении существенные успехи достигнуты усилиями исследователей в области лингвистики в целом и когнитивной лингвистики в первую очередь. Информацию о представленных в тексте фактах лингвисты обычно связывают с его содержанием, вводя в рассмотрение семантическое пространство, которое наделяют определенной структурой [3]. Этот факт, как и многие другие результаты лингвистических исследований могут составить фундамент, на котором целесообразно разрабатывать методы формализации текстов, причем в самых различных сферах деятельности людей.

Объектом изучения лингвистики является язык, который самым тесным образом связан со всеми аспектами мыслительной и познавательной деятельности человека. Поэтому лингвисты тесно взаимодействуют с представителями различных наук и во все большей мере информатики [4]. В рамках рассматриваемой тематики, следует обратить внимание на работу [5], где описан подход, согласно которому формирование и использование ассоциативно-онтологического представления данных обеспечивает получение средств автоматического реферирования научных и технических документов. Особо следует отметить работы, проведенные в рамках когнитивной информатики, которые связаны с изучением отображения в информационных системах процессов эволюции знаковых систем, в частности, построения взаимно-однозначных отношений между концептами и компьютерными кодами [6].

Несмотря на существенные успехи в области содержательной обработке текстов программно-техническими комплексами, это направление остается актуальным и исследования здесь активно продолжаются.

Представляется перспективной разработка средств поддержки работы с учебными материалами, базирующихся на изучении структуры представленных в учебных текстах знаний, и вот почему. Систему как таковую можно рассматривать в виде совокупности неких объектов (элементов системы) связанных между собой в интересах достижения определенной цели. При этом следует обратить внимание на то, что структура, отображая связи между элементами, несет в себе существенный объем сведений о данной системе (как говорят, имеет большую информативность). В этой связи при

моделировании системы кроме сведений о ее элементах непременно следует учитывать связи между последними [7].

Учебный текст, представляющий собой зафиксированную информацию, как средство обучения, ориентированное на усвоение обучаемым соответствующих знаний, умений, навыков (далее термин «знание» используем в обобщенном смысле), обычно имеет четко выраженную структуру, которая несет существенно важные сведения о содержании, организации и реализации учебных процессов. Тем самым учебный текст – объект нашего рассмотрения – может рассматриваться как система.

Обладание сведениями о структуре учебного материала, представленными в формальном виде обеспечивает возможность разработки моделей и алгоритмов, а на их основе построения программно-технических комплексов, позволяющих более эффективно решать целый ряд педагогических задач. Центральным на этом пути шагом является, по нашему мнению, построение формальной модели структуры изложенных в учебных материалах знаний, представленных в виде текстов, написанных на естественном языке.

Итак, ставится задача: для данного представленного на естественном языке учебного текста сформировать представленную в формальном виде его знаниевую структуру.

Построение формальной модели структуры знаний представленного на естественном языке текста

Учебный текст, как правило, рубрицирован, представляя собой последовательность отдельных вложенных друг в друга частей (рублик). Он может быть построен, например, из нескольких разделов, каждый из которых состоит из глав, которые в свою очередь разбиты на параграфы, последние на пункты и т.д. Фрагмент текста в каждой из этих рубрик посвящен изложению определенной и относительно самостоятельной части осваиваемого материала. Обычно процесс обучения и реализуется в соответствии с порядком следования рубрик [8].

Вместе с тем даже самая небольшая из рубрицированных частей текста представляет собой весьма сложную конструкцию, составленную из еще более мелких фрагментов. Принципиально важным является возможность выделения самых малых (с педагогической точки зрения) представленных в данном тексте фрагментов. Такую порцию подлежащей усвоению логически замкнутой информации, обычно получающей при введении свое имя и воспринимаемую в дальнейшем как неделимый фрагмент знаний, будем называть учебным элементом. К таковым относим: определения сущностей и описание их свойств, отдельные действия, формулы, правила и др. Важно

отметить, что каждый учебный элемент является носителем уникальной в рамках данного текста информации.

После усвоения учебного элемента последний должен восприниматься обучаемым как носящая атомарный характер целостность, т.е. при упоминании его имени у него в сознании должна актуализироваться информация, усвоенная при знакомстве с соответствующим учебным элементом (что это, как образуется, как используется, ...).

Для удобства понимания последующего изложения, рассмотренные и обработанные (представленные в формализованном виде и включенные в модель) учебные элементы будем называть квантами знаний. При таком понимании кванты знаний служат «строительным материалом» для построения новых (последующих) подлежащих усвоению учебных элементов. Для этого между несколькими определенными квантами знаний (между введенными ранее и вводимым в данный момент) фиксируются описанные в тексте соответствующие связи. Тем самым вся совокупность представленных в учебном тексте квантов знаний и множество связей между ними образуют определенную знаниевую структуру, которая является предметом нашего рассмотрения.

Процесс изучения учебного материала обычно и построен в виде определенной последовательности процедур освоения новых учебных элементов. При этом могут даваться некоторые определения и описываться связи между несколькими учебными элементами, относительно которых предполагается, что они были введены ранее.

Например, учебный элемент, в котором дается определение обратной матрицы, может быть введен следующим текстом: обратной относительно некоторой исходной квадратной матрицы A называется такая матрица A^{-1} , произведение которой на матрицу A образует единичную матрицу E , а далее следует формула: $A \cdot A^{-1} = E$.

Здесь для введения учебного элемента «обратная матрица» непосредственно используются учебные элементы: «квадратная матрица», «единичная матрица» и «умножение матриц», относительно которых предполагается, что они были освоены обучаемым ранее. Отметим, что часть учебных элементов не вводится в рассматриваемом материале, поскольку предполагается, что обучаемый знаком с ними к началу изучения данного материала (будем называть таковые исходными учебными элементами). Так для студента школьные знания, как правило, полагаются исходными.

Таким образом, учебный текст можно трактовать, как знаниевую систему, базовым множеством которой является совокупность исходных и вводимых учебных элементов,

а также связи между ними, что образует содержательную структуру этого текста. (напомним, латинское слово *textus* означает ткань, соединение).

Построение модели структуры знаний предлагается осуществлять в ходе пошагового конструктивного процесса, каждый шаг которого заключается в выявлении и фиксации в формальной форме очередного учебного элемента и его связей с другими элементами. Этот процесс можно интерпретировать как формирование ориентированного графа, представляющего собой формальную (математическую) модель структуры представленных в тексте знаний.

Изложим процесс построения модели более детально. Пусть по мере прочтения рассматриваемого текста на шаге k находится описание очередного учебного элемента. Обозначим его через y_k . Ему будет соотноситься выходная вершина формируемого подграфа, отображающего описание этого учебного элемента.

В представленном в тексте описании y_k непосредственно используется некоторое множество, введенных, как правило, ранее, учебных элементов, обозначим это множество через $X_k = \{x_i\}_k$. Следовательно, формируемый на k -м шаге ориентированный подграф $\Delta G(k)$ имеет вид

$$\Delta G(k) = (y_k, \{x_i\}_k, \{i, k\} \mid \forall i < k),$$

где x_i – вершины, из которых дуги $\{x_{ik}\}$ выходят и все они входят в вершину y_k .

Примечание. Если в $\{x_i\}_k$ отсутствует учебный элемент x_i , это означает его отсутствие в рассматриваемом тексте и, как следствие в построенной части графа. Такие ситуации свидетельствуют о наличии брака в учебном тексте, о чем делаются соответствующие пометки. Формально это приводит к наличию в графе так называемых «пустых» вершин.

Здесь каждая дуга (i, k) играет чисто формальную роль ($\forall x_i \in X_k: x_{ik} = 1$), поскольку отображает следующий факт: учебный элемент x_i , введенный когда-то на i -м шаге используется на k -м шаге при введении элемента y_k . Дело в том, что само наличие элемента x_i в $\Delta G(k)$ уже свидетельствует об этом.

Положим, что исходные для данного учебного материала элементы считаются введенными на нулевом шаге $X_0 = \{x_{0i}\}$. На практике же удобнее принимать решение о том, является ли данный учебный элемент исходным или нет, на каждом шаге обработки текста.

Для последующего применения формируемой модели все описания учебных элементов удобно фиксировать в виде соответствующих фреймов. В каждом фрейме, представляющем элементу u_k , помимо перечисления всех использованных при его введении элементов и связей между ними, целесообразно фиксировать собственно фрагмент текста, где описано, как соотносятся между собой эти элементы. В дальнейшем это позволит существенно расширять возможности применения модели.

Сформированный в результате k -го шага подграф $\Delta G(k)$ объединяют с уже построенным за все предыдущие шаги графом $G(k-1)$, т.е. происходит наращивание формируемой модели

$$G(k-1) \cup \Delta G(k) = G(k)$$

и переходят к рассмотрению следующего по тексту учебного элемента.

Если весь текст рассмотрен, то в итоге имеем модель структуры знаний, содержащихся в рассмотренном учебном тексте

$$G(n) = \langle \{x_i\}, \{i, j\} \rangle, \forall i \in \overline{0, n}, j \in \overline{1, n}, \forall i < j,$$

где n - количество введенных в рассматриваемом тексте учебных единиц.

Перспективы применения модели структуры знаний учебного текста

Наличие математической модели структуры представленных в учебном материале знаний позволит более эффективно решать в сфере образования ряд прикладных задач. В первую очередь это задачи анализа учебных текстов, например, определение количества введенных учебных элементов, выявление нарушений логики изложения, фиксация не введенных (или введенных не вовремя) учебных элементов и др., что позволяет строить объективные автоматически реализуемые методики оценки качества учебных материалов и их сравнения.

Изложенный подход может быть применен к построению модели структуры знаний всей совокупности учебных дисциплин образовательной программы в целом или ее предметной части, например, математической подготовки студентов. Это позволяет решать вопросы целостности и полноты программы, согласования очередности изучаемых дисциплин и их частей, а также вопросы оптимизации содержательного наполнения программы.

Развиваемые системы активной поддержки самоподготовки обучающихся, благодаря использованию в них модели структуры знаний изучаемого учебного материала, могут приобретать новые качества [9]. Это, в частности, возможность автоматизировать процессы поиска «истоков незнания» конкретного пункта программы и формирования логики их устранения, повышения уровня мотивации обучающегося к освоению материала должным образом, сделать более удобными процедуры перманентного и объективного контроля текущего состояния подготовленности обучающегося и, что очень важно, за счет получения количественных характеристик протекания процесса обучения повысить уровень объективности оценки этого состояния.

Пусть обучаемый работает в автоматизированной обучающей среде, которая включает в свой арсенал и модель структуры представленных в изучаемом материале знаний. Это обстоятельство позволяет для различных педагогических ситуаций формировать соответствующие траектории движения обучающегося в пространстве его знаний, строить в складывающейся учебной ситуации целесообразные траектории перемещения в этом пространстве и рекомендовать обучающему его последующие действия, устанавливать значения дифференциальных и интегральных оценок уровня его подготовки. Поскольку все действия обучающегося могут быть зафиксированы, и эта информация обработана, то могут быть получены рекомендации для преподавателя, разработчиков программ и авторов учебных материалов.

Таким образом, наличие в системах системы активной поддержки самоподготовки обучающихся подсистем, базирующихся на модели структуры знаний учебного материала, обеспечивает возможность в режиме реального времени контролировать и анализировать все действия обучающегося, выявлять и идентифицировать складывающиеся педагогические ситуации, формировать адекватные им траектории последующих действий и целенаправленно вести по ней обучающегося.

Список библиографических ссылок (на языке оригинала)

1. Лебедева М.Ю., Веселовская Т.С., Купрещенко О.Ф. Особенности восприятия и понимания цифровых текстов: междисциплинарный взгляд. Перспективы Науки и Образования. Международный электронный научный журнал. 2020;4. URL: rpojjournal.wordpress.com/archive20/20-04.

2. Бедарев Н.В., Войнов А.А. Тексты на естественном языке и методы извлечения структурированных данных // Международная научно-технологическая конференция

студентов и молодых ученых «Молодёжь. Наука. Технологии». Новосибирск. 2018; т. 2. С. 37-42.

3. Pavlovskiy I.S. The semantic models of large terminological texts // Tenth International Conference Management of Large-Scale System Development (MLSD). 2017. P. 1–5.

4. Gaspers Judith, Cimiano Philipp, Rohlfing Katharina, Wrede Britta. Constructing a Language From Scratch: Combining Bottom–Up and Top–Down Learning Processes in a Computational Model of Language Acquisition. *IEEE Transactions on Cognitive and Developmental Systems*. 2017; 9(2):183–196.

5. Кулешов С.В., Зайцева А.А., Марков В.С. Ассоциативно-онтологический подход к обработке текстов на естественном языке. *Интеллектуальные технологии на транспорте*. 2015; 4: 40–45

6. Зацман И.М. Семиотическая модель взаимосвязей концептов, информационных объектов и компьютерных кодов. *Информатика и ее применение*. 2009; 3(2):65–81.

7. Воронов М.В., Пименов В.И. Формализация регулятивных текстов. *Информатика и автоматизация*. 2021; 20(3):562–590.

8. Сабинаина А.А. Учебный текст: структура и прагматика. *Известия РГПУ им. А. И. Герцена*. 2009; 97:222-225.

9. Воронов М.В. Герасименко П.В., Письменский Г.И. Разработка виртуальной среды самостоятельной подготовки студентов. *Инновации в образовании*. 2017; 9:31–44.

References (на английском языке)

1. Lebedeva M.YU., Veselovskaya T.S., Kupreshchenko O.F. Osobennosti vospriyatiya i ponimaniya tsifrovyykh tekstov: mezhdistsiplinarnyy vzglyad. *Perspektivy Nauki i Obrazovaniya. Mezhdunarodnyy elektronnyy nauchnyy zhurnal*. 2020;4. [Electronic resource]. Available at: pnojurnal.wordpress.com/archive20/20-04. (In Russian)

2. Bedarev N.V., Voynov A.A. Teksty na yestestvennom yazyke i metody izvlecheniya strukturirovannykh dannykh // *Mezhdunarodnaya nauchno-tekhnologicheskaya konferentsiya studentov i molodykh uchenykh «Molodozh'. Nauka. Tekhnologii»*. Novosibirsk. 2018; т. 2. P. 37-42. (In Russian)

3. Pavlovskiy I.S. The semantic models of large terminological texts // Tenth International Conference Management of Large-Scale System Development (MLSD). 2017. P. 1–5.

4. Gaspers Judith, Cimiano Philipp, Rohlfing Katharina, Wrede Britta. Constructing a Language From Scratch: Combining Bottom–Up and Top–Down Learning Processes in a Computational Model of Language Acquisition. *IEEE Transactions on Cognitive and*

Developmental Systems [IEEE Transactions on Cognitive and Developmental Systems]. 2017; 9(2):183–196.

5. Kuleshov S.V., Zaytseva A.A., Markov V.S. Assotsiativno-ontologicheskij podkhod k obrabotke tekstov na yestestvennom yazyke. *Intellektual'nyye tekhnologii na transporte*. 2015; 4: 40–45. (In Russian)

6. Zatsman I.M. Semioticheskaya model' vzaimosvyazey kontseptov, informatsionnykh ob"yektov i komp'yuternykh kodov. *Informatika i yeye primeneniye*. 2009; 3(2):65–81.

7. Voronov M.V., Pimenov V.I. Formalizatsiya regul'yativnykh tekstov. *Informatika i avtomatizatsiya* [Informatics and automation]. 2021; 20(3):562–590. (In Russian)

8. Sabinina A.A. Uchebnyy tekst: struktura i pragmatika. *Izvestiya RGPU im. A. I. Gertsena*. 2009; 97:222-225. (In Russian)

9. Voronov M.V. Gerasimenko P.V., Pis'menskiy G.I. Razrabotka virtual'noy sredy samostoyatel'noy podgotovki studentov. *Innovatsii v obrazovanii*. 2017; 9:31–44. (In Russian)