

Мир науки. Педагогика и психология / World of Science. Pedagogy and psychology <https://mir-nauki.com>

2022, №2, Том 10 / 2022, No 2, Vol 10 <https://mir-nauki.com/issue-2-2022.html>

URL статьи: <https://mir-nauki.com/PDF/46PDMN222.pdf>

**Ссылка для цитирования этой статьи:**

Каверзина, Д. Е. Структурирование контента электронного обучающего курса на основе семантического анализа глоссария / Д. Е. Каверзина, Ю. В. Вайнштейн // Мир науки. Педагогика и психология. — 2022. — Т. 10. — № 2. — URL: <https://mir-nauki.com/PDF/46PDMN222.pdf>

**For citation:**

Kaverzina D.E., Vainshtein Yu.V. Structuring the content of an e-learning course based on the semantic analysis of the glossary. *World of Science. Pedagogy and psychology*, 10(2): 46PDMN222. Available at: <https://mir-nauki.com/PDF/46PDMN222.pdf>. (In Russ., abstract in Eng.).

УДК 004.81+372.851

**Каверзина Дарья Евгеньевна**

ФГАОУ ВО «Сибирский федеральный университет», Красноярск, Россия  
Ассистент кафедры «Прикладной математики и компьютерной безопасности»  
E-mail: [dkaverzina@sfu-kras.ru](mailto:dkaverzina@sfu-kras.ru)

**Вайнштейн Юлия Владимировна**

ФГАОУ ВО «Сибирский федеральный университет», Красноярск, Россия  
Доцент кафедры «Прикладной математики и компьютерной безопасности»  
Кандидат технических наук, доцент  
E-mail: [yweinstein@sfu-kras.ru](mailto:yweinstein@sfu-kras.ru); [julia\\_ww@mail.ru](mailto:julia_ww@mail.ru)  
ORCID: <https://orcid.org/0000-0002-8370-7970>  
РИНЦ: [https://elibrary.ru/author\\_profile.asp?id=165831](https://elibrary.ru/author_profile.asp?id=165831)

## Структурирование контента электронного обучающего курса на основе семантического анализа глоссария

**Аннотация.** В условиях интенсивного внедрения цифровых технологий в образовательный процесс и цифровой трансформации образования активно развиваются подходы и методики онлайн обучения, интеграции онлайн и офлайн обучения, а также разрабатывается учебно-методическое обеспечение в электронной информационно-образовательной среде вуза. Лавинообразное нарастание объемов электронного образовательного контента актуализирует проблему его структурирования на основе создания новых подходов, включающих современные методы анализа и формализации данных.

В статье предложен подход к формализованному структурированию электронного обучающего контента на основе автоматизированного анализа глоссария с применением семантической модели данных. Особенностью предложенной семантической модели данных выступает включение в нее понятий предметной области дисциплины, индексируемых посредством их семантического содержания. В условиях возрастания объемов учебной деятельности, самостоятельно осуществляемой обучающимися в электронной среде, логически связанная структура предметного содержания ЭОК приобретает особую значимость. Преимуществом семантической модели данных перед другими моделями представления знаний выступает связывание близких понятий в единый пучок сети через установление связей между понятиями образовательного контента. Раскрыты особенности разработанного авторского программного обеспечения анализа глоссария, обеспечивающего автоматизированное построение иерархической структуры образовательного контента — дерева понятий.

Осуществлена апробация предложенного подхода при разработке электронного обучающего курса по дисциплине «Моделирование и формализация» для студентов направления 09.03.03 — «Прикладная информатика» Института космических и информационных технологий Сибирского федерального университета. Педагогический эксперимент подтвердил результативность образовательного процесса с применением электронного обучающего курса, контент которого структурирован на основе автоматизированного семантического анализа глоссария. Обозначены достоинства и перспективы развития предложенного подхода.

**Ключевые слова:** образовательный контент; дерево понятий; глоссарий электронного обучающего курса; онлайн обучение; семантическая модель данных

### Введение

В условиях цифровой трансформации образования наряду с интенсивным внедрением цифровых технологий в образовательный процесс активно развиваются подходы и методики онлайн обучения, интеграции онлайн и офлайн обучения, а также разрабатывается учебно-методическое обеспечение в электронной информационно-образовательной среде вуза. В условиях распространения новой коронавирусной инфекции образовательным организациям высшего и среднего образования пришлось осуществить экстренный переход на дистанционное обучение, эффективным средством которого стали электронные обучающие курсы (ЭОК). Но если раньше они использовались преподавателями преимущественно в условиях веб-поддержки обучения, то сегодня форматы их применения значительно расширились. Пандемия наглядно продемонстрировала необходимость развития электронного обучения и важность разработки инновационных электронных обучающих курсов и ресурсов. Лавинообразное увеличение количества электронных обучающих курсов и ресурсов и нарастание объемов электронного образовательного контента актуализирует проблему его структурирования на основе создания новых подходов, включающих современные методы анализа и формализации данных.

Активное применение в учебном процессе электронного образовательного контента существенно повышает требования к его структуре и содержанию. Наибольшее распространение получило представление образовательного контента электронных обучающих курсов в соответствии с принципами микрообучения [1–3]. Стоит заметить, что, несмотря на интенсивное применение микрообучения, существующие практики в основном ориентированы на разделение учебного материала на фрагменты и основным принципом разделения является объем материала и предполагаемое время его изучения. Такой подход зачастую не включает перестройку последовательности изучения учебного материала и влечет фрагментарность и отсутствие логической связности разрабатываемых ЭОК. Вместе с тем к основным критериям построения структуры образовательного контента ЭОК можно отнести критерии целостности, логической связанности, полноты и проверяемости [2]. Можно утверждать, что структурирование контента предметной области дисциплины представляет собой одну из проблем при создании ЭОК. Логически связанная структура предметного содержания ЭОК приобретает особую значимость в онлайн обучении в связи с возрастанием объемов самостоятельно осуществляемой обучающимися учебной деятельности.

Все вышесказанное обуславливает поиск новых подходов к проектированию структуры электронного обучающего курса на основе применения и развития современных методов анализа и формализации данных.

## Методы

При выявлении теоретических оснований и особенностей построения электронных обучающих курсов, ресурсов и структурирования образовательного контента онлайн обучения применялся метод сравнительно-сопоставительного анализа психолого-педагогической, научно-методической литературы. Для получения обратной связи о формируемых образовательных результатах нами использовались такие эмпирические методы, как анкетирование, тестирование, наблюдение обучающихся в электронной среде. Методологическую основу исследования составляет системный подход, рассматривающий обучение как целостный процесс во взаимосвязи всех его компонент и интегрирующий преимущества компетентностного, личностно-ориентированного, деятельностного, предметно-информационного и средового подходов в образовательном процессе. Теоретическую основу исследования составили работы в области теории информатизации образования, методики электронного и дистанционного обучения, структурирования образовательного контента и другие [4–6].

## Результаты

Существуют различные подходы к структурированию образовательного контента предметного обучения [7–11]. В статье мы предлагаем подход к проектированию структуры образовательного контента электронного обучающего курса на основе семантической модели данных, особенностью которой выступает включение в нее понятий, индексируемых посредством их семантического содержания [12]. Отметим, что под образовательным контентом мы будем понимать системное, логически связанное предметное содержание электронного обучающего курса. К элементам образовательного контента можно отнести видеоматериалы, аудиозаписи, текстовые учебные материалы, задания, тесты, опросы и др. Преимуществом применяемой нами семантической модели данных перед другими моделями представления знаний выступает связывание близких понятий в единый пучок сети, то есть установления связей между понятиями образовательного контента [13]. В результате структурирования контента предметной области дисциплины осуществляется построение иерархической структуры *дерева понятий*. При этом вершины дерева представляют собой понятия, а ребра связи между ними. Все понятия дерева группируются по уровням, на каждом из которых — понятия, которые получены в результате одинакового числа применений операции деления к исходному понятию (рис. 1).

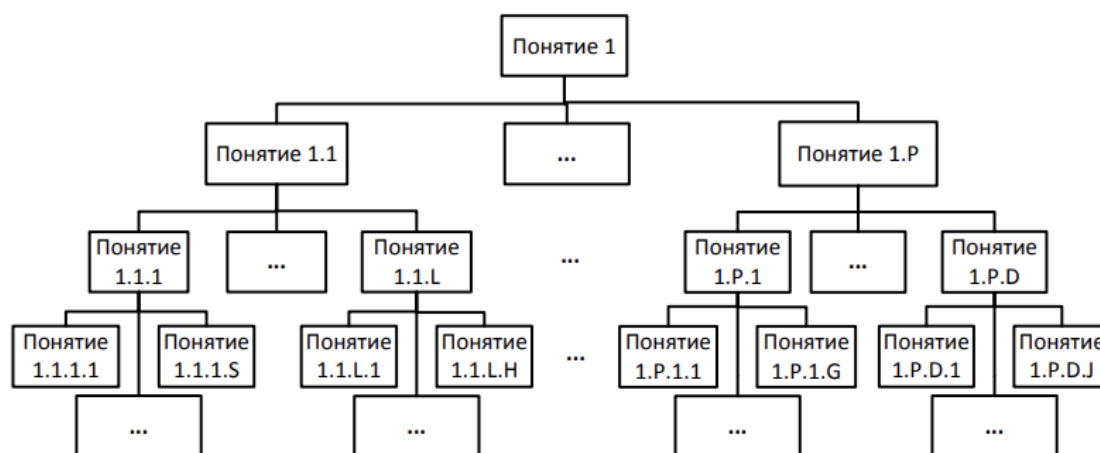


Рисунок 1. Дерево понятий образовательного контента (составлено/разработано автором)

Представление контента ЭОК в виде семантической сети позволяет структурировать дисциплину на уровне основных понятий и заложить основу для основных учебных действий: усвоения понятий в сфере их определения, выявления основных признаков и свойств изучаемых объектов и выявления структурно-логических связей в рамках изучаемой теории [14]. Представление предметной области в виде дерева отражает существующее на множестве понятий отношение частичного порядка по включению объемов понятий и включает ближайшие отношения иерархии между понятиями: «родовидовые», «часть-целое» и не отражает транзитивные [15; 16].

Сравнить относительный вклад различных понятий и их связей в общий образовательный контент ЭОК, выявить взаимосвязи между понятиями, определить порядок изучения образовательного контента и формализовать построение дерева понятий мы предлагаем на основе анализа смысловых числовых характеристик, таких как вес понятия и вес связи понятий. Формализованное построение дерева понятий представляет собой процесс выделения понятий нулевого, первого, второго и последующего уровней. При этом к нулевому уровню дерева понятий мы относим основное понятие, не зависящее от других. К понятиям первого уровня — понятия, которые находятся в подчинении у основного понятия и при этом не связаны между собой. К понятиям второго и последующего уровней — понятия, которые связаны ровно с одним понятием предыдущего уровня и не связаны с другими. Такой подход позволяет иерархически упорядочить их, при этом, на нулевом уровне дерева располагается наиболее общее фундаментальное понятие, затем на первом уровне оно конкретизируется через базовые понятия, которые на втором уровне конкретизируются через ключевые понятия и далее — конкретизируются вплоть до частных понятий. То есть реализуется иерархическая цепочка: классы понятий — обобщенные понятия — конкретные (элементарные) понятия предметной области дисциплины. Осуществляя обход дерева понятий, например, методом поиска в ширину мы получаем последовательность изучения понятий дисциплины, то есть осуществляем структурирование учебного материала электронного обучающего курса.

Предлагаемый нами подход к построению последовательности изучения учебных материалов позволил нам выделить следующие основные этапы разработки ЭОК:

- определение образовательных результатов дисциплины;
- определение всех необходимых понятий предметной области и формирование глоссария;
- структурирование предметной области дисциплины в виде дерева понятий;
- определение последовательности изучения понятий ЭОК;
- разработка контента ЭОК согласно полученной структуре.

Структурирование предметной области дисциплины в виде дерева понятий представляет собой достаточно трудоемкий процесс для преподавателя. Поэтому особый интерес в процессе создания ЭОК представляет программное обеспечение, предназначенное для автоматизированного построения деревьев понятий [17]. Распространение получило программное обеспечение, предназначенное для анализа содержания разработанных ЭОК, то есть выделения понятий, их весов и связей между ними из текстового учебного материала. Мы предлагаем подход к структурированию содержания ЭОК на основе автоматизированного анализа глоссария, который строго структурирован. Все связи и отношения исследуемых объектов глоссария даются в системном виде, допускающем исключительно однозначное толкование и при этом каждое определение находится в логической взаимосвязи с другими терминами и понятиями.

Рассмотрим определение количественных характеристик понятий и их связей. Отметим, что под *понятием* мы будем понимать некоторое представление о классе предметов или явлений. *Термин* — это слово или словосочетание, выражающее специальное понятие, которое соотносено с другими понятиями в данной предметной области и имеющее уникальное научное определение. *Определение* — это текстовое описание, раскрывающее содержание понятия или термина.

Вес понятия мы находим следующим образом:

$$W_j = \frac{k_j}{k} + \frac{m_j}{m},$$

где

$W_j$  — вес понятия  $j$ ;

$k_j$  — количество вхождений терминов в определение понятия  $j$ ;

$k$  — общее количество вхождений всех терминов в определения понятий глоссария;

$m_j$  — количество всех определений понятий, содержащих  $j$ -ый термин;

$m$  — общее количество определений понятий во всем глоссарии.

При этом первое слагаемое оценивает вес определения понятия  $j$ , второе слагаемое оценивает вес термина  $j$ , а их сумма — это и есть полный вес понятия.

Оценка веса понятия  $j$  (мера «значимости» или «важности» понятия) находится по формуле:  $\overline{W}_j = \frac{w_j}{\max_j w_j} * 100$ . Значение смыслового веса варьируется от 1 до 100 и отражает важность понятия в учебном материале. Максимальное значение, равное 100, говорит о том, что понятие является ключевым. Значение, близкое к единице показывает, что понятие является редко используемым в учебном материале.

Вес связи понятия  $j$  мы находим следующим образом:

$$Vp_j = kpv_j * kps_j,$$

где  $Vp_j$  — вес связи понятия  $j$ ;

$kpv_j$  — количество терминов, входящих в определение понятия  $j$ ;

$kps_j$  — количество терминов, содержащих понятие  $j$ .

Оценка веса связи понятия  $j$  (мера «значимости» или «важности» связи понятия с остальными):  $\overline{Vp}_j = \frac{vp_j}{\max_j vp_j} * 100$ . Она показывает, как связано анализируемое понятие с остальными понятиями и учитывает:

- количество подчиненных понятий;
- количество родительских понятий.

Вес связи варьируется от 0 до 100. Если понятие имеет одностороннюю связь, т. е. употребляется в определениях других понятий, а в своем определении не имеет других терминов или наоборот, то вес такой связи нулевой. Данное понятие не является связующим звеном между уровнями дерева понятий, наглядным доказательством чего является построенное с помощью ИС дерево. Главное слово также имеет вес связи равный 0, так как не имеет родовидовых связей и является корнем дерева. Если же понятие имеет вес связи равный 100, то данное понятие имеет максимальное число подчиненных и родительских понятий.

Разработанное авторское программное обеспечение осуществляет автоматизированное построение семантической модели на основе анализа глоссария предметной области дисциплины [18]. Уникальностью разработанного программного модуля выступает включение в алгоритмы помимо механизма выделения терминов, механизмов и методов оценки веса понятий и связи, что развивает алгоритмы, существующего программного обеспечения, например, «LanA-Key» [19]. Программный модуль анализа глоссариев электронных обучающих ресурсов и курсов авторов статьи выполняет следующие функции:

1. Осуществляет семантический разбор глоссария (поиск терминов понятий с определениями и терминов, находящихся внутри определений других понятий).
2. Осуществляет поиск главного понятия — корня дерева понятий.
3. Анализирует, полученную по средствам разбора глоссария, информацию:
  - вычисляет вес понятия и вес его связи с другими понятиями;
  - определяет редко используемые в глоссарии понятия и понятия, не имеющие определения.
4. Формирует таблицу весов понятий и их связей для наглядного представления результатов анализа.
5. Формирует модель образовательного контента ЭОК в виде дерева понятий.

Апробация предложенного авторами подхода к структурированию контента предметной области осуществлена при разработке электронных обучающих курсов для студентов направления 09.03.03 — «Прикладная информатика» Сибирского федерального университета. Рассмотрим в качестве примера структурирование образовательного контента по дисциплине «Моделирование и формализация». На рисунке 2 представлено дерево понятий, построенное по результатам анализа глоссария.

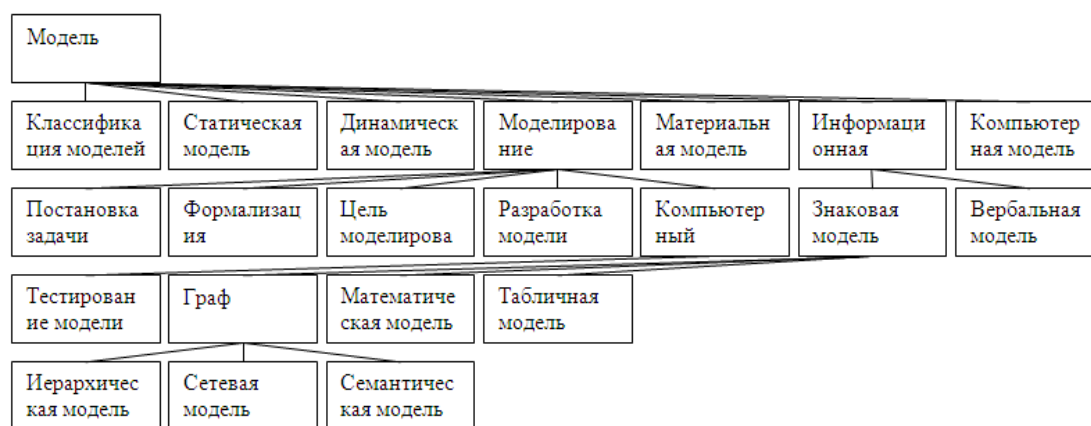


Рисунок 2. Дерево понятий, построенное ПО (составлено/разработано автором)

Каждое из понятий рассматривается как одна из потенциальных единиц образовательного контента ЭОК, что помогает сформировать структуру предметного содержания следующим образом:

1. **Модель**
  - 1.1. Классификация моделей
  - 1.2. Материальные модели
  - 1.3. Статические модели

- 1.4. Динамические модели
- 1.5. Компьютерные модели
- 2. Информационные модели**
  - 2.1. *Знаковые модели*
    - 2.1.1. Табличные модели
    - 2.1.2. Математические модели
    - 2.1.3. *Графы*
      - 2.1.3.1. Иерархические модели
      - 2.1.3.2. Сетевые модели
      - 2.1.3.3. Семантические модели
  - 2.2. Вербальные модели
- 3. Моделирование**
  - 3.1. Формализация
  - 3.2. Постановка задачи
  - 3.3. Цель моделирования
  - 3.4. Разработка модели
  - 3.5. *Компьютерный эксперимент*
    - 3.5.1. Тестирование модели

Представление предметной области в виде дерева понятий позволило структурировать дисциплину на уровне основных понятий и заложить основу для основных учебных действий обучающихся: изучения понятий и структурно-логических связей между ними, осуществления операций над понятиями предметной области и др. Педагогический эксперимент подтвердил результативность образовательного процесса с применением ЭОК «Моделирование и формализация», структурирование образовательного контента которого было осуществлено на основе построения семантической модели предметной области.

### **Вывод**

Проектирование структуры электронного обучающего курса на основе автоматизированного анализа глоссария обладает универсальностью с точки зрения его применения в учебном процессе по любой дисциплине и позволяет получать логически обоснованные последовательности изучения образовательного контента.

Предложенный подход автоматизированного анализа глоссария дает возможность определять не только структуру электронного обучающего курса (разделы, темы, параграфы), но и последовательность введения понятий, более подробно представлять наиболее важные из них или исключать понятия и связанный с ними учебный материал с низким смысловым весом. Это может помочь преподавателям профессиональных образовательных организаций и образовательных организаций высшего образования, учителям общеобразовательных организаций в процессе создания ими ЭОК определить логически обоснованную последовательность изучения учебного материала. Мы полагаем, что структурирование контента электронного обучающего курса на основе семантического анализа глоссария

существенно снизит трудоемкость педагогов и разработчиков при создании электронных обучающих курсов и повысит их вовлеченность в этот процесс.

## ЛИТЕРАТУРА

1. Федосеев А.А. К вопросу об уменьшении объема порций учебного материала при электронном обучении / А.А. Федосеев // Информатика и её применение. — 2016. — Том 10. — Выпуск 3. — С. 105–110.
2. Dolasinski M.J., Reynolds J. Microlearning: A New Learning Model // Journal of Hospitality & Tourism Research. — 2020. — № 44(3). P. 551–561.
3. Fernandez J. The microlearning trend: Accommodating cultural and cognitive. — URL: <https://learningsolutionsmag.com/articles/1578/the-microlearning-trend-accommodating-cultural-and-cognitive-shifts> (дата обращения: 27.04.2022).
4. Пак Н.И. Концепция трансформационных и перевернутых электронных учебников / Н.И. Пак, Е.Г. Потупчик, Л.Б. Хегай // Вестник Российского университета дружбы народов. Серия: Информатизация образования. — 2020. — Т. 17. — № 2. — С. 153–168.
5. Стариченко Б.Е. Цифровизация образования: реалии и проблемы // Педагогическое образование в России. — 2020. — № 4. — С. 16–26.
6. Уваров А.Ю. От компьютеризации до цифровой трансформации образования // Информатика и образование. — 2019. — № 4(303). — С. 5–11.
7. Атанов Г.А. Моделирование учебной предметной области, или предметная модель обучаемого // Образовательные технологии и общество. — 2000. — № 3(3). — С. 111–124.
8. Белякова Е.Г. Взаимодействие студентов вуза с образовательным контентом в условиях информационной образовательной среды / Е.Г. Белякова, И.Г. Захарова // Образование и наука. — 2019. — Т. 21. — № 3. — С. 77–105.
9. Бова В.В. К вопросу о построении инфологической модели представления образовательного контента в интегрированной научно-образовательной среде // Известия Южного федерального университета. Технические науки. — 2007. — Выпуск 77. — № 2. — С. 200–206.
10. Голикова Е.А. Опыт построения структуры курса «Дискретная математика» с отслеживанием логических связей между ее компонентами // Современные проблемы науки и образования. — 2018. — № 5. — URL: <https://science-education.ru/ru/article/view?id=28165> (дата обращения: 27.04.2022).
11. Тельнов Ю.Ф. Принципы и методы семантического структурирования информационно-образовательного пространства на основе реализации онтологического подхода // Статистика и экономика. — 2014. — № 1. — С. 187–191.
12. Вайнштейн Ю.В., Каверзина Д.Е. Проектирование структуры электронного обучающего курса на основе автоматизированного анализа глоссария. — URL: <http://nesinmis.ru/Vainshtein-Yu-V-Kaverzina-D-E/> (дата обращения: 27.04.2022).



13. Захарова И.Г., Семикин В.А. О некоторых общих принципах разработки обучающих систем // Математическое и информационное моделирование: Сборник статей. — Тюмень: Изд-во ТюмГУ, 2000. — С. 171–177.
14. Куваев М.Р. Методика преподавания математики в вузе. — Томск: изд-во Томского университета, 1990. — 387 с.
15. Вайнштейн Ю.В., Есин Р.В., Цибульский Г.М. Модель образовательного контента: от структурирования понятий к адаптивному обучению // Открытое образование. — 2021. — № 2. — С. 44–52.
16. Воробьев Г.А. Модель структуризации содержания обучения на основе построения иерархической понятийной сети знаний / Г.А. Воробьев, П.А. Корнев, В.Н. Малыш // Вестник РУДН. — 2009. — № 4. — С. 94–101.
17. Ботов Д.С., Кленин Ю.Д. Применение алгоритмов автоматического извлечения ключевых слов и моделей векторного представления слов в задаче анализа образовательного контента // Труды Международной научной конференции Научное электронное издание. — 2017. — С. 31–36.
18. Программный модуль анализа глоссариев электронных обучающих ресурсов и курсов: свидетельство о регистрации программы для ЭВМ 2022617472 Российская Федерация / Вайнштейн Ю.В., Каверзина Д.Е., заявка № 2022616829 от 14.04.2022.
19. LanA Consulting Aps. — URL: <http://www.lanaconsult.com/products.htm/> (дата обращения: 27.04.2022).

**Kaverzina Daria Evgenievna**

Siberian Federal University, Krasnoyarsk, Russia  
E-mail: [dkaverzina@sfu-kras.ru](mailto:dkaverzina@sfu-kras.ru)

**Vainshtein Yulia Vladimirovna**

Siberian Federal University, Krasnoyarsk, Russia  
E-mail: [yweinstein@sfu-kras.ru](mailto:yweinstein@sfu-kras.ru); [julia\\_ww@mail.ru](mailto:julia_ww@mail.ru)  
ORCID: <https://orcid.org/0000-0002-8370-7970>  
RSCI: [https://elibrary.ru/author\\_profile.asp?id=165831](https://elibrary.ru/author_profile.asp?id=165831)

## Structuring the content of an e-learning course based on the semantic analysis of the glossary

**Abstract.** In the context of the intensive introduction of digital technologies in the educational process and the digital transformation of education, approaches and methods of online learning, integration of online and offline learning are actively developing, as well as educational and methodological support is being developed in the electronic information and educational environment of the university. An avalanche-like increase in the volume of electronic educational content actualizes the problem of its structuring based on the creation of new approaches, including modern methods of data analysis and formalization.

The article proposes an approach to the formalized structuring of e-learning content based on automated glossary analysis using a semantic data model. A feature of the proposed semantic data model is the inclusion in it of the concepts of the subject area of the discipline, indexed by means of their semantic content. In the context of an increase in the volume of educational activities independently carried out by students in an electronic environment, the logically connected structure of the subject content of ELC acquires special significance. The advantage of the semantic data model over other models of knowledge representation is the linking of related concepts into a single bundle of the network through the establishment of links between the concepts of educational content. The features of the developed author's software for the analysis of the glossary, which provides automated construction of the hierarchical structure of educational content — a tree of concepts, are disclosed.

The proposed approach was tested in the development of an electronic training course in the discipline "Modeling and Formalization" for students of the direction 09.03.03 — "Applied Informatics" of the Institute of Space and Information Technologies of the Siberian Federal University. The pedagogical experiment confirmed the effectiveness of the educational process using an e-learning course, the content of which is structured based on an automated semantic analysis of the glossary. The advantages and prospects for the development of the proposed approach are outlined.

**Keywords:** educational content; concept tree; e-learning course glossary; online learning; semantic data model