

## УЧЁТ СТРУКТУРЫ ДОКУМЕНТА В МЕТОДЕ АВТОМАТИЧЕСКОГО АННОТИРОВАНИЯ МАТЕМАТИЧЕСКИХ ПОНЯТИЙ В ОБРАЗОВАТЕЛЬНЫХ ТЕКСТАХ

К. С. Николаев<sup>[0000-0003-3204-238X]</sup>

*Лаборатория суперкомпьютерного моделирования Отделения суперкомпьютерных систем и параллельных вычислений ФГБУ «Национальный исследовательский центр «Курчатовский институт»», г. Москва;  
Казанский (Приволжский) федеральный университет, г. Казань.*

konnikolaeff@yandex.ru

### **Аннотация**

Обогащение образовательных текстов семантическим содержимым (в частности, дополнение документа гиперссылками на страницы сервиса, отображающего подробную информацию о понятиях, используемых в тексте) способствует повышению эффективности усвоения материала обучающимися. Существующие методы семантической разметки образовательных текстов не учитывают структурные особенности таких документов, что приводит к избыточному распознаванию понятий.

В статье описано развитие метода автоматического аннотирования математических понятий в образовательных математических текстах путем добавления функционала для учета структуры образовательного документа. Основное назначение метода заключается в обработке образовательных материалов курса дистанционного образования «Технология решения планиметрических задач». Соблюдение единого шаблона при создании страниц курса позволяет применить анализ веб-разметки страниц и ключевых слов, примененных создателями курса. Основной задачей в данном процессе является определение типа ячеек таблицы, в которых находятся текстовые фрагменты образовательных материалов. В соответствии с рекомендациями создателей курса, определения необходимо выделять в ячейках, содержащих постановку задачи, а также в тех блоках, где указаны входные данные задачи. Определение типа ячеек таблиц производится с помо-

щью анализа их атрибутов и поиска ключевых слов в их содержимом. Такое ограничение распознаваемых фрагментов текста позволяет улучшить восприятие страниц курса учеником и повысить качество усвоения учебного материала.

***Ключевые слова:** семантический анализ, математическая онтология, дидактические отношения, математическое образование, разметка документа.*

## **ВВЕДЕНИЕ**

Применение онтологий в образовательных приложениях является актуальной задачей, так как применение классических образовательных материалов без семантического содержимого не позволяет в полной мере вовлечь учащегося в образовательный процесс и снижает общую эффективность обучения. Также онтологии позволяют студенту сформировать более цельную картину изучаемой дисциплины, изучить отношения и закономерности между отдельными понятиями в проблемной области.

Онтологии и семантические технологии часто применяются в задачах улучшения образовательного процесса. В работе [1] авторы предложили онтологию для поддержки и унификации процесса составления учебной программы. В работе [2] применены технологии Semantic Web для персонализации обратной связи в образовательных средах. В [3] использована методология MOnto для разработки многоязычных приложений в образовательной среде. Онтологии также активно применяются в обучении анатомии и медицины в целом. В [4] представлен подход к персонализации программ образовательных курсов с помощью онтологии. В [5] онтологии и семантические технологии применены для корректировки и персонализации назначения заданий студентам. Авторы [6] предложили использовать онтологии в качестве основы рекомендационной системы.

Для использования преимуществ, предлагаемых семантическими технологиями при обработке и представлении математического контента, была предложена экосистема OntoMath<sup>Edu</sup> [7]. Она содержит набор инструментов для применения в задачах математического образования.

Основным источником понятий и отношений между понятиями названной экосистемы служит онтология OntoMath<sup>Edu</sup> [8–18]. Она подробно отражает предметную область планиметрии и предназначена для включения в программу курса по планиметрии, который разрабатывается в Казанском федеральном университете для общеобразовательных средних школ. В онтологии OntoMath<sup>Edu</sup> выделены три уровня:

1. Предметный уровень, включающий понятия, которые не зависят от языка и входят в программу обучения в средней школе.
2. Лингвистический уровень, содержащий текстовые представления понятий предметного уровня онтологии.
3. Уровень онтологии верхнего уровня, предоставляющий метаонтологические аннотации.

Отличительной особенностью онтологии OntoMath<sup>Edu</sup> от других предметных онтологий является наличие дидактических отношений между понятиями: образовательные системы, образовательные уровни и пререквизиты [19–21]. Образовательные системы отражают различия в подходе к изучению понятий планиметрии в различных странах [22] (на данный момент в онтологии представлены образовательные системы РФ, Великобритании и США). Образовательные уровни позволяют разделить множество понятий на подмножества, содержащие понятия, изучаемые на том или ином образовательном уровне (например, в образовательной системе РФ для программы планиметрии: 7 класс, 8 класс, профильный 8 класс и т. д.). Пререквизиты отражают последовательность изучения понятий в соответствии с образовательной системой [23].

В состав экосистемы OntoMath<sup>Edu</sup> входит набор взаимосвязанных сервисов, обеспечивающих обработку, генерацию и визуализацию математических документов [24, 25]. Этот набор содержит сервисы, приведенные ниже (см. также рис. 1).

1. Сервис обогащения онтологии, формирующий набор рекомендаций для экспертов по улучшению структуры онтологии путем поиска контекстуально близких понятий, между которыми отсутствуют прямые связи в онтологии.
  2. Сервис генерации тестовых заданий, формирующий тестовые задания с вводом ответа и интерактивные задания по восстановлению иерархии понятий образовательной онтологии [26].
-

3. Сервис семантического аннотирования формул в pdf-документах, формирующий семантическое представление встроенных формул на основе включенных в них компонентов для формирования коллекции документов, применяемых в семантическом поисковике формул [27].
4. Сервис выделения математических понятий в документах.
5. Сервис визуализации подграфов семантических сетей [28].

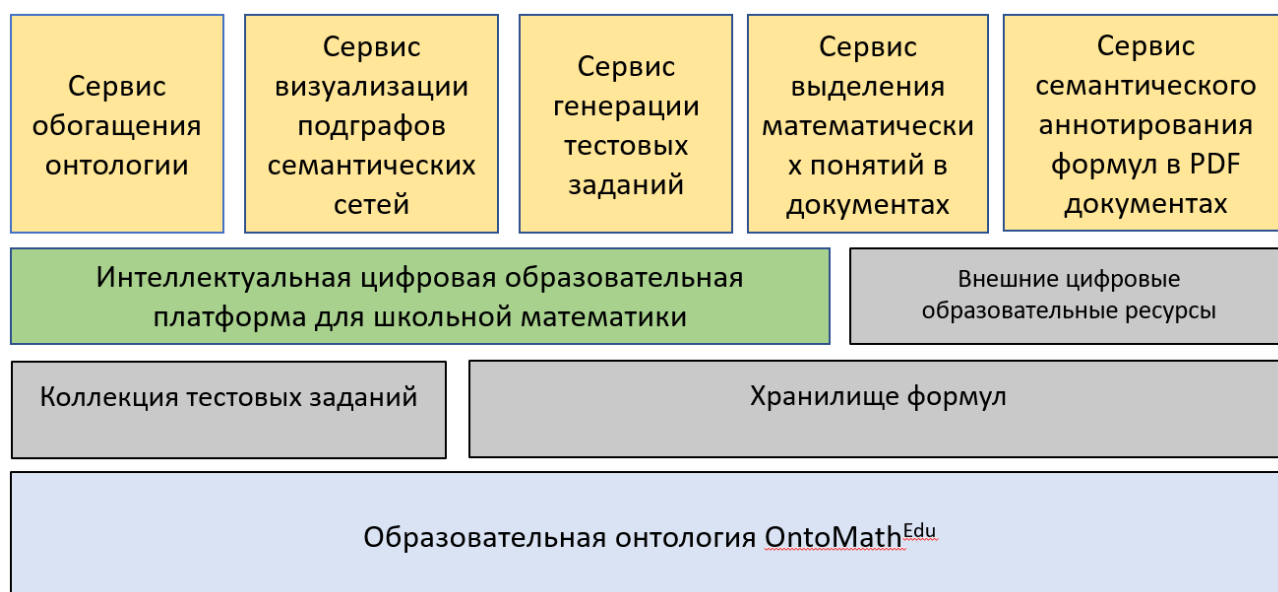


Рис. 1. Структура экосистемы OntoMath<sup>Edu</sup>

Помимо перечисленных сервисов в состав экосистемы входят наборы данных, необходимые для работы сервисов и являющиеся результатами работы сервисов. Хранилище формул содержит набор математических утверждений, представленных на естественном языке, в нотации LaTeX и в формализации OpenMath [29]. Коллекция тестовых заданий содержит множество таких заданий, сформированных на основе математических утверждений в формате OpenMath из хранилища формул, а также множество интерактивных заданий на восстановление фрагментов иерархии понятий онтологии OntoMath<sup>Edu</sup> (<https://github.com/CLLKazan/OntoMathEdu>). Ссылки на внешние образовательные ресурсы интегрированы в концепты онтологии и содержат информацию о концептах на других ресурсах в интернете.

Основой сервиса выделения математических понятий в документах является метод автоматического аннотирования математических понятий в образовательных математических текстах. Сервис осуществляет поиск математических понятий из онтологии OntoMath<sup>Edu</sup> в образовательных текстах в формате HTML и производит обогащение документа ссылками на сервис визуализации подграфов семантических сетей, отображающий подробную информацию о математическом понятии и понятиях, связанных с ним. В следующем разделе описаны метод автоматического аннотирования математических понятий в образовательных математических текстах и его улучшение с помощью учета разметки документов.

### **МЕТОД АВТОМАТИЧЕСКОГО АННОТИРОВАНИЯ МАТЕМАТИЧЕСКИХ ПОНЯТИЙ В ОБРАЗОВАТЕЛЬНЫХ МАТЕМАТИЧЕСКИХ ТЕКСТАХ**

Блок-схема метода автоматического аннотирования математических понятий в образовательных текстах представлена на рис. 2. Далее перечислены основные шаги указанного метода.

1. Предобработка документа, включающая разбор HTML-документа на теги, извлечение текстового содержимого, нормализация всех слов в предложениях.

2. Выбор значимых слов (существительные, прилагательные в полной и краткой формах) в пределах каждого предложения. Формирование цепочек значимых слов с длинами от 1 до N, где N – количество значимых слов в предложении.

3. Предобработка концептов онтологии OntoMath<sup>Edu</sup>: нормализация всех слов в названиях концептов (например, концепт «Вершина многоугольника» преобразуется в «вершина многоугольник») и выделение значимых слов.

4. Сравнение цепочек слов в предложениях с множеством значимых слов в концептах онтологии, фильтрация и сохранение результатов разметки.

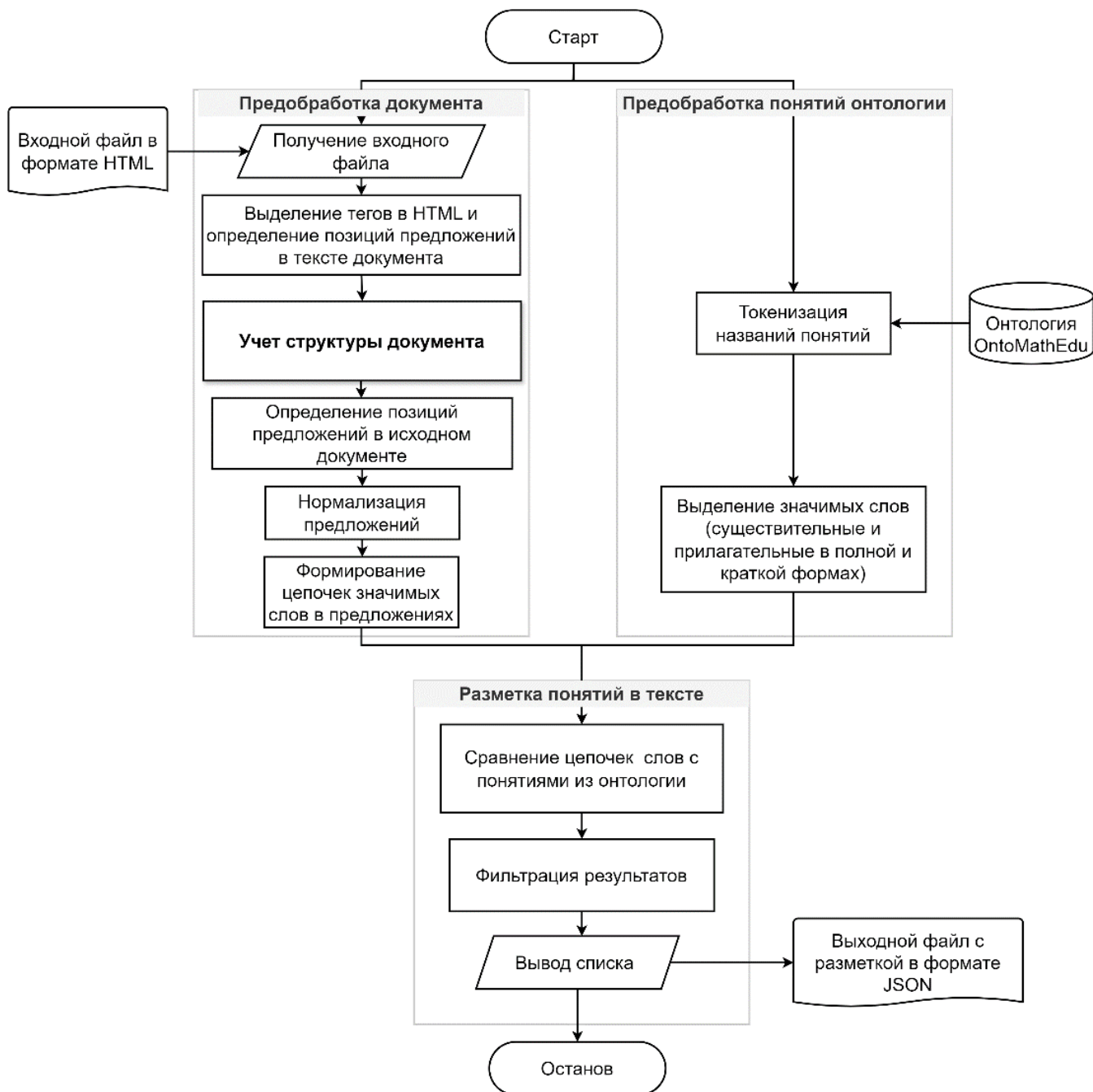


Рис. 2. Блок-схема метода автоматического аннотирования математических понятий в образовательных текстах

При полном совпадении сочетания слов и названия концепта каждому слову в сочетании присваивается ссылка на концепт онтологии. Если слово оказывается связанным с двумя и более концептами, приоритет отдается понятию с наибольшим количеством слов. Например, в тексте «В прямоугольном треугольнике проведена высота из угла А» на слово «треугольнике» будут претендовать

понятия «Треугольник» и «Прямоугольный треугольник», и в результате понятие «Треугольник» будет игнорироваться.

Для управления процессом аннотирования и редактирования результатов было разработано программное обеспечение, позволяющее в интерактивном режиме производить разметку документов в формате HTML с визуализацией текущего состояния документа и производить привязку концептов онтологии OntoMath<sup>Edu</sup> к фрагментам текста документа. На рис. 3 приведен интерфейс этого программного обеспечения.

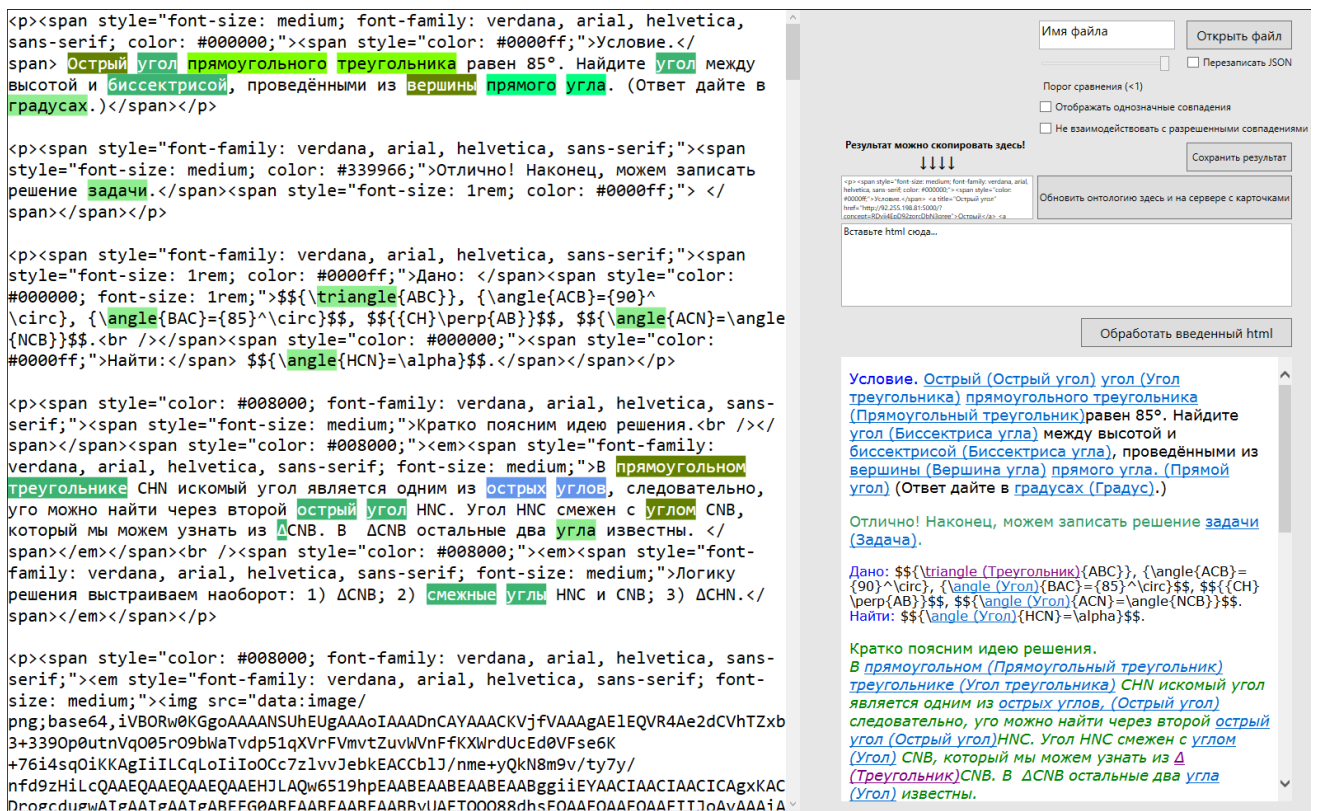


Рис. 3. Интерфейс программного обеспечения для управления методом аннотирования математических понятий в образовательных текстах

Названное программное обеспечение применено при разработке материалов курса дистанционного образования «Технология решения планиметрических задач». С учетом отзывов авторов курса возникла необходимость улучшения процесса аннотации математических понятий с помощью учета структуры образовательных документов для исключения фрагментов документа из процесса аннотации. Модификация указанного программного обеспечения описана в следующем разделе.

## УЧЁТ ОГРАНИЧЕНИЙ В МЕТОДЕ АВТОМАТИЧЕСКОГО АННОТИРОВАНИЯ МАТЕМАТИЧЕСКИХ ПОНЯТИЙ В ОБРАЗОВАТЕЛЬНЫХ МАТЕМАТИЧЕСКИХ ТЕКСТАХ

Учебные материалы дистанционного курса «Технологии решения планиметрических задач» представлены в виде таблицы, в отдельных ячейках которых размещены условие задачи, вспомогательная информация и блоки с вариантами заданий. В процессе аннотирования должны участвовать только ячейки с условием задачи и блоки с вариантами заданий. На рис. 4 приведен фрагмент HTML-разметки, содержащей условие задачи. Жирным шрифтом выделены теги, на основе которых происходит фильтрация ячеек таблицы (`<table>` `<tbody>` `<td>`, производится поиск самой широкой ячейки). Кроме того, на этом рисунке жирным шрифтом выделен пример ключевой фразы, по которой дополнительно фильтруются ячейки таблицы («Условие задачи», «Дано»). Набор ключевых фраз будет расширяться в зависимости от разнообразия форматов образовательных документов в дистанционном курсе.

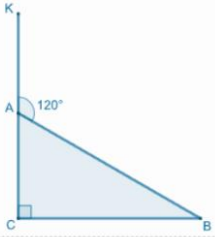


```
<table border="0">
  <tbody>
    <tr>
      <td colspan="3">
        <p><span
            style="color: #ff6600;"><strong>Условие задачи</strong>
        </span>В <a title="Прямоугольный треугольник"
          href="http://84.39.245.172:5000">прямоугольном
          треугольнике</a> <em>ABC</em> с <a title="Прямой угол"
          href="http://84.39.245.172:5000">прямым
          углом</a> <em>C</em> <a
            title="Внешний угол треугольника"
            href="http://84.39.245.172:5000">внешний
            угол</a> при вершине <em>A</em> равен  $120^\circ$ , <em>AC
          </em>+ <em>AB</em> = 18. Найти <em>AC</em> и
          <em>AB</em>.</p>
        </td>
      </tr>
    </tbody>
  </table>
```

Рис. 4. Результат аннотирования математических понятий без учета структуры документа

На рис. 5 представлено сравнение результатов автоматического аннотирования до и после добавления учета разметки документа. Как видно из рис. 5 а, без учета разметки метод аннотирования производит разметку также во вспомогательных блоках (фрагменты текста с фиолетовым цветом шрифта). После исключения блоков, не содержащих ключевые фразы, метод аннотирования затрагивает только значимые блоки (рис. 5 б).

**Условие задачи.** В прямоугольном треугольнике  $ABC$  с прямым углом  $C$  внешний угол при вершине  $A$  равен  $120^\circ$ ,  $AC + AB = 18$ . Найти  $AC$  и  $AB$ .



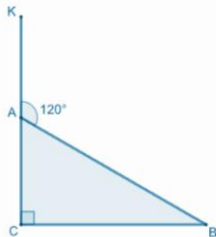
На продолжении стороны  $AC$  отметим точку  $K$  и составим краткую запись.

Кстати, давай договоримся обозначать углы одной буквой там, где угол определен однозначно, например, углы при вершинах  $B$  и  $C$  будем обозначать  $\angle B$  и  $\angle C$ . При вершине  $A$  имеются два угла, их мы обозначим тремя буквами –  $\angle BAC$  и  $\angle KAB$ .

Давай проверим, сможешь ли ты распознать краткую запись, соответствующую требованиям.

<b>Ано:</b> $\Delta ABC$ , $\angle C = 90^\circ$ , $\angle KAB$ – внешний с $\angle A = 120^\circ$ , $AC + AB = 18$ см.	<b>Ано:</b> $\Delta ABC$ , $\angle C = 90^\circ$ , $\angle KAB$ – внешний с $\angle BAC = 120^\circ$ , $AC + AB = 18$ см.	<b>Ано:</b> $\Delta ABC$ , $\angle C = 90^\circ$ , $\angle A$ – внешний с $\angle BAC = 120^\circ$ , $AC + AB = 18$ см.
<b>Н</b> <b>йти:</b> $AC, AB$	<b>Н</b> <b>йти:</b> $AC, AB$	<b>Н</b> <b>йти:</b> $AC, AB$

**Условие задачи.** В прямоугольном треугольнике  $ABC$  с прямым углом  $C$  внешний угол при вершине  $A$  равен  $120^\circ$ ,  $AC + AB = 18$ . Найти  $AC$  и  $AB$ .



На продолжении стороны  $AC$  отметим точку  $K$  и составим краткую запись.

Кстати, давай договоримся обозначать углы одной буквой там, где угол определен однозначно, например, углы при вершинах  $B$  и  $C$  будем обозначать  $\angle B$  и  $\angle C$ . При вершине  $A$  имеются два угла, их мы обозначим тремя буквами –  $\angle BAC$  и  $\angle KAB$ .

Давай проверим, сможешь ли ты распознать краткую запись, соответствующую требованиям.

<b>Ано:</b> $\Delta ABC$ , $\angle C = 90^\circ$ , $\angle KAB$ – внешний с $\angle A = 120^\circ$ , $AC + AB = 18$ см.	<b>Ано:</b> $\Delta ABC$ , $\angle C = 90^\circ$ , $\angle KAB$ – внешний с $\angle BAC = 120^\circ$ , $AC + AB = 18$ см.	<b>Ано:</b> $\Delta ABC$ , $\angle C = 90^\circ$ , $\angle A$ – внешний с $\angle BAC = 120^\circ$ , $AC + AB = 18$ см.
<b>Н</b> <b>йти:</b> $AC, AB$	<b>Н</b> <b>йти:</b> $AC, AB$	<b>Н</b> <b>йти:</b> $AC, AB$

**а)**
**б)**

Рис. 5. Результат аннотирования математических понятий. а) без учета структуры документа, б) с учетом структуры документа

## ЗАКЛЮЧЕНИЕ

Представлено развитие метода автоматического аннотирования математических понятий в образовательных текстах с помощью учета структуры документов в формате HTML, позволяющего исключить из разметки фрагменты текста, содержащие вспомогательную информацию, и уменьшить объем ручной работы для авторов дистанционного курса «Технология решения планиметрических задач». В будущем планируется улучшить учет разметки образовательных документов.

Укажем основные направления планируемого расширения функциональных возможностей разметки документов.

1. Условия фильтрации фрагментов документа, представленные выше, специфичны для стиля оформления учебных материалов в конкретном дистанционном курсе. Необходима разработка более универсальных параметров фильтрации блоков документа, подходящих для большого количества учебных материалов.

2. Расширение функционала для настройки анализа разметки документа в программное обеспечение для визуализации и редактирования разметки.

### **Благодарности**

Статья подготовлена в рамках государственного задания Федерального государственного бюджетного учреждения «Национальный исследовательский центр «Курчатовский институт» (тема FNEF-2024-0014).

### **СПИСОК ЛИТЕРАТУРЫ**

1. *Zouri M., Ferworn A.* An Ontology-Based Approach for Curriculum Mapping in Higher Education // 2021 IEEE 11th Annu. Comput. Commun. Work. Conf. 2021. P. 141–147.

2. *Cheniti Belcadhi L.* Personalized feedback for self-assessment in lifelong learning environments based on semantic web // Comput. Human Behav. 2016. Vol. 55. P. 562–570.

3. *Florrence M.* Building a multilingual ontology for education domain using monto method // Comput. Sci. Inf. Technol. 2020.

4. *Ibrahim M.E., Yang Y., Ndzi D.L., Yang G., Al-Maliki M.* Ontology-Based Personalized Course Recommendation Framework // IEEE Access. 2019. Vol. 7. P. 5180–5199.

5. *Radovic M., Petrovic N., Tosic M.* An Ontology-Driven Learning Assessment Using the Script Concordance Test // Appl. Sci. 2022. Vol. 12, No. 3.

6. *Obeid C., Lahoud I., El Khoury H., Champin P.A.* Ontology-based Recommender System in Higher Education // Web Conf. 2018 – Companion World Wide Web Conf. WWW 2018. 2018. P. 1031–1034.

7. *Nikolaev K.S., Nevzorova O.A., Falileeva M.V.* Developing the OntoMathEdu ecosystem for educational applications // CEUR Workshop Proc. 2021. Vol. 2910. P. 81–87.

8. *Невзорова О.А., Фалилеева М.В., Липачёв Е.К., Шакирова Л.Р., Дюпина А.Э., Кириллович А.В.* Образовательная онтология OntoMathEdu: проблемы онтологического инжиниринга // XX национальная конференция по искусственному интеллекту с международным участием, КИИ-2022. Труды конференции. В 2 т. Том 1. М.: Издательство МЭИ, 2022. С. 177–188.

9. *Nevzorova O.A., Falileeva M.V., Kirillovich A.V., Lipachev E.K., Shakirova L.R., Dyupina A.E.* OntoMathEdu Educational Ontology: Problems of Ontological Engineering // *Pattern Recognit. Image Anal.* 2023. Vol. 33. P. 460–466.

<https://doi.org/10.1134/S1054661823030367>

10. *Dyupina A., Falileeva M.* Problems of Designing Polylingual Ontology OntoMathEdu // *Communications in Computer and Information Science.* 2020. Vol. 1242. P. 179–190. [https://doi.org/10.1007/978-3-030-65218-0\\_14](https://doi.org/10.1007/978-3-030-65218-0_14)

11. *Shakirova L., Falileeva M., Kirillovich A., Lipachev E., Nevzorova O., Nevzorov V.* Modeling and Evaluation of the Mathematical Educational Ontology // *SSI-2019 Scientific Services & Internet Proceedings of the 21st Conference on Scientific Services & Internet (SSI-2019) Novorossiysk-Abrau, Russia, September 23–28, 2019.* CEUR Workshop Proceedings. 2019. Vol. 2543. P. 305–319.

12. *Шакирова Л.Р., Фалилеева М.В., Кириллович А.В., Липачёв Е.К., Невзорова О.А., Невзоров В.Н.* Образовательная математическая онтология OntoMathEdu: структура и отношения // *Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23–28 сентября 2019 г., г. Новороссийск).* М.: ИПМ им. М.В. Келдыша. 2019. С. 653–661.

13. *Nevzorova O., Shakirova L., Falileeva M., Kirillovich A., Lipachev E., Nevzorov V.* OntoMathEdu Educational Mathematical Ontology: Annotation of Concepts // *Selected Contributions of the "Russian Advances in Artificial Intelligence" Track at RCAI 2020 co located with 18th Russian Conference on Artificial Intelligence (RCAI 2020).* Moscow, Russia, October 10–16, 2020. Vol. 2648. P. 157–168.

14. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L.* OntoMathEdu: A Linguistically Grounded Educational Mathematical Ontology // *Intelligent Computer Mathematics. Series: Lecture Notes in Computer Science.* 2020. Vol. 12236. P. 157–172. <https://doi.org/10.1007/978-3-030-53518-6>

15. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L.* OntoMathEdu: Towards an Educational Mathematical Ontology // *Edwin Brady et al. (Eds.). Workshop Papers at 12th Conference on Intelligent Computer Mathematics (ICM-WS 2019), Prague, Czech Republic, 8–12 July 2019.* CEUR Workshop Proceedings. 2020. Vol. 2634.

16. *Shakirova L., Falileeva M., Kirillovich A., Lipachev E., Nevzorova O., Nevzorov V.* Modeling and Evaluation of the Mathematical Educational Ontology // Proceedings of the 21st Conference on Scientific Services & Internet (SSI-2019). CEUR Workshop Proceedings. Vol. 2543. P. 305–319.

17. *Шакирова Л.Р., Фалилеева М.В., Кириллович А.В., Липачев Е.К., Невзорова О.А., Невзоров В.Н.* Образовательная математическая онтология OntoMathEdu: структура и отношения // Научный сервис в сети Интернет. 2019. Том 21. С. 653–661. <https://doi.org/10.20948/abrau-2019-84>

18. *Шакирова Л.Р., Фалилеева М.В., Кириллович А.В., Невзорова О.А.* Проектирование образовательной математической онтологии: проблемы и методы решения на примере курса планиметрии // Труды международной конференции по компьютерной и когнитивной лингвистике: Материалы XV международной конференции по компьютерной и когнитивной лингвистике TEL-2018 в 2-х томах, Казань. 2018. С. 393–405. EDN SQGRVY.

19. *Falileeva M., Kirillovich A., Shakirova L., Nevzorova O., Lipachev E., Dyupina A.* OntoMathEdu Educational Mathematical Ontology: Prerequisites, Educational Levels and Educational Projections // CEUR Workshop Proceedings. 2020. Vol. 2784. P. 346–351.

20. *Фалилеева М.В., Кириллович А.В., Невзорова О.А., Шакирова Л.Р., Липачёв Е.К., Дюпина А.Э.* Системы образовательных проекций, уровней и пререквизитов математической онтологии OntoMathEdu // Электронные библиотеки. 2021. Т. 24. № 3. С. 505–530. <https://doi.org/10.26907/1562-5419-2021-24-3-505-530>

21. *Nevzorova O.A., Falileeva M.V., Kirillovich A.V., Shakirova L.R., Lipachev E.K., Dyupina A.E.* Modeling of didactic relationships in the OntoMathEdu educational mathematical ontology // CEUR Workshop Proceedings. 2021. Vol. 2910. P. 11–21.

22. *Фалилеева М.В., Кириллович А.В., Невзорова О.А., Шакирова Л.Р., Липачёв Е.К., Дюпина А.Э.* Разработка образовательных проекций математической онтологии OntoMathEdu // Научный сервис в сети Интернет. 2020. Т. 22. С. 604–611. <https://doi.org/10.20948/abrau-2020-26>. EDN WBNWIC

23. *Kirillovich A., Falileeva M., Nevzorova O., Lipachev E., Dyupina A., Shakirova L.* Prerequisite Relationships of the OntoMathEdu Educational Mathematical On-

tology // J.C. Figueroa-García et al. (Eds.). Proceedings of the 8th Workshop on Engineering Applications (WEA 2021). Communications in Computer and Information Science. 2021. Vol. 1431. P. 517–524. [https://doi.org/10.1007/978-3-030-86702-7\\_44](https://doi.org/10.1007/978-3-030-86702-7_44)

24. *Nevzorova O., Shakirova L., Falileeva M., Nikolaev K.* Towards Ontology-Based Software Services for Teaching School Mathematics // L. Gómez Chova et al. (Eds.). Proceedings of the 15th International Technology, Education and Development Conference (INTED 2021), March 8–9th. 2021. P. 5288–5297. <https://doi.org/10.21125/inted.2021.1085>

25. *Невзорова О.А., Липачёв Е.К., Николаев К.С.* Семантические сервисы цифровой экосистемы OntoMath для математического образования // *Электронные библиотеки*. 2023. Т. 26. № 4. С. 538–564. <https://doi.org/10.26907/1562-5419-2023-26-4-538-569>

26. *Nikolaev K.S., Kirillovich A.V., Nevzorova O.A.* A corpus-based approach to elementary geometry knowledge test generation. // INTED2020 Proceedings. 2020. P. 6342–6348. <https://doi.org/10.21125/inted.2020.1710>

27. *Невзорова О.А., Николаев К.С.* Семантическое аннотирование математических формул в PDF-документах // *Электронные библиотеки*. 2022. Т. 25. № 6. С. 616–639. <https://doi.org/10.26907/1562-5419-2022-25-6-616-639>

28. *Николаев К.С.* Сервис генерации учебных карточек математических понятий для дистанционного курса по геометрии // *Электронные библиотеки*. 2023. Т. 26. № 3. P. 365–377. <https://doi.org/10.26907/1562-5419-2023-26-3-365-377>

29. *Nevzorova O.A., Kirillovich A.V., Nikolaev K.S., Galiaskarova K.R.* Towards a Semantically Annotated Corpus of Educational Mathematical Texts in Russian (short paper) // International Conference on Data Analytics and Management in Data Intensive Domains. 2019. P. 299–305. CEUR Workshop Proceedings. 2019. Vol. 2523. P. 299–305.

## TAKING INTO ACCOUNT THE STRUCTURE OF THE DOCUMENT IN THE METHOD OF AUTOMATIC ANNOTATION OF MATHEMATICAL CONCEPTS IN EDUCATIONAL TEXTS

K. S. Nikolaev<sup>[0000-0003-3204-238X]</sup>

*Supercomputer Modeling Laboratory of the Department of Supercomputer Systems and Parallel Computing of the Federal State Budgetary Institution "National Research Center "Kurchatov Institute"", Moscow;  
Kazan (Volga Region) Federal University, Kazan.*

konnikolaeff@yandex.ru

### **Abstract**

The enrichment of educational texts with semantic content (in particular, adding hyperlinks to the pages of the service that displays detailed information about concepts in the text) helps to increase the efficiency of students' assimilation of the material. The existing methods of semantic markup of educational texts do not take into account the structural features of such documents, which leads to excessive recognition of concepts. This article describes the development of the method of automatic annotation of mathematical concepts in educational mathematical texts by adding functionality to account for the structure of an educational document. The main purpose of the method is to process educational materials of the distance education course "Technology for solving planimetric problems". Following a single template when creating course pages allows you to apply an analysis of the web page markup and keywords used by the course creators. The main task in this process is to determine the type of table cell containing text fragments of educational materials. In accordance with the recommendations of the course creators, definitions should be highlighted in the cells containing the task statement, as well as in those blocks where the input data of the task is indicated. The type of table cells is determined by analyzing their attributes and searching for keywords in their contents. This limitation of recognizable text fragments will improve the student's perception of the course pages and improve the quality of learning.

**Keywords:** *semantic analysis, mathematical ontology, didactic relations, mathematical education, document markup.*

## REFERENCES

1. *Zouri M., Ferworn A.* An Ontology-Based Approach for Curriculum Mapping in Higher Education // 2021 IEEE 11th Annu. Comput. Commun. Work. Conf. 2021. P. 141–147.
2. *Cheniti Belcadhi L.* Personalized feedback for self-assessment in lifelong learning environments based on semantic web // Comput. Human Behav. 2016. Vol. 55. P. 562–570.
3. *Florrence M.* Building a multilingual ontology for education domain using monto method // Comput. Sci. Inf. Technol. 2020.
4. *Ibrahim M.E., Yang Y., Ndzi D. L., Yang G., Al-Maliki M.* Ontology-Based Personalized Course Recommendation Framework // IEEE Access. 2019. Vol. 7. P. 5180–5199.
5. *Radovic M., Petrovic N., Tosic M.* An Ontology-Driven Learning Assessment Using the Script Concordance Test // Appl. Sci. 2022. Vol. 12, No. 3.
6. *Obeid C., Lahoud I., El Khoury H., Champin P. A.* Ontology-based Recommender System in Higher Education // Web Conf. 2018 – Companion World Wide Web Conf. WWW 2018. 2018. P. 1031–1034.
7. *Nikolaev K.S., Nevzorova O.A., Falileeva M.V.* Developing the OntoMathEdu ecosystem for educational applications // CEUR Workshop Proc. 2021. Vol. 2910. P. 81–87.
8. *Nevzorova O.A., Falileeva M.V., Lipachev E.K., Shakirova L.R., Dyupina A.E., Kirillovich A.V.* Obrazovatel'naya ontologiya OntoMathEdu: problemy` ontologicheskogo inzhiniringa // XX nacional'naya konferenciya po iskusstvennomu intellektu s mezhdunarodny`m uchastiem, KII-2022. Trudy` konferencii. V 2 t. Tom 1. M.: Izdatel'stvo ME`I, 2022. S. 177–188.
9. *Nevzorova O.A., Falileeva M.V., Kirillovich A.V., Lipachev E.K., Shakirova L.R., Dyupina A.E.* OntoMathEdu Educational Ontology: Problems of Ontological Engineering // Pattern Recognit. Image Anal. 2023. Vol. 33. P. 460–466.  
<https://doi.org/10.1134/S1054661823030367>
10. *Dyupina A., Falileeva M.* Problems of Designing Polylingual Ontology OntoMathEdu // Communications in Computer and Information Science. 2020. Vol. 1242. P. 179–190. [https://doi.org/10.1007/978-3-030-65218-0\\_14](https://doi.org/10.1007/978-3-030-65218-0_14)



11. *Shakirova L., Falileeva M., Kirillovich A., Lipachev E., Nevzorova O., Nevzorov V.* Modeling and Evaluation of the Mathematical Educational Ontology // SSI-2019 Scientific Services & Internet Proceedings of the 21st Conference on Scientific Services & Internet (SSI-2019) Novorossiysk-Abrau, Russia, September 23–28, 2019. CEUR Workshop Proceedings. 2019. Vol. 2543. P. 305–319.

12. *Shakirova L.R., Falileeva M.V., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Nevzorov V.N.* Obrazovatel'naya matematicheskaya ontologiya OntoMathEdu: struktura i otnosheniya // Nauchny`j servis v seti Internet: trudy` XXI Vserossijskoj nauchnoj konferencii (23–28 sentyabrya 2019 g., g. Novorossiysk). M.: IPM im. M.V. Keldy`sha. 2019. S. 653–661.

13. *Nevzorova O., Shakirova L., Falileeva M., Kirillovich A., Nevzorov V., Lipachev E.* OntoMathEdu Educational Mathematical Ontology: Annotation of Concepts // Selected Contributions of the "Russian Advances in Artificial Intelligence" Track at RCAI 2020 co-located with 18th Russian Conference on Artificial Intelligence (RCAI 2020). Moscow, Russia, October 10–16, 2020. Vol. 2648. P. 157–168.

14. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L.* OntoMathEdu: A Linguistically Grounded Educational Mathematical Ontology // Intelligent Computer Mathematics. Series: Lecture Notes in Computer Science. 2020. Vol. 12236. P. 157–172. <https://doi.org/10.1007/978-3-030-53518-6>

15. *Kirillovich A., Nevzorova O., Falileeva M., Lipachev E., Shakirova L.* OntoMathEdu: Towards an Educational Mathematical Ontology // Edwin Brady et al. (Eds.). Workshop Papers at 12th Conference on Intelligent Computer Mathematics (CICM-WS 2019), Prague, Czech Republic, 8–12 July 2019. CEUR Workshop Proceedings. 2020. Vol. 2634.

16. *Shakirova L., Falileeva M., Kirillovich A., Lipachev E., Nevzorova O., Nevzorov V.* Modeling and Evaluation of the Mathematical Educational Ontology // Proceedings of the 21st Conference on Scientific Services & Internet (SSI-2019). CEUR Workshop Proceedings. Vol. 2543. P. 305–319.

17. *Shakirova L.R., Falileeva M.V., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Nevzorov V.N.* Obrazovatel'naya matematicheskaya ontologiya OntoMathEdu: struktura i otnosheniya // Nauchny`j servis v seti Internet. 2019. № 21. S. 653–661. <https://doi.org/10.20948/abrau-2019-84>

18. *Shakirova L.R., Falileeva M.V., Kirillovich A.V., Nevzorova O.A.* Proektirovanie obrazovatel'noj matematicheskoy ontologii: problemy` i metody` resheniya na primere kursa planimetrii // Trudy` mezhdunarodnoj konferencii po komp'yuternoj i kognitivnoj lingvistike: Materialy` XV mezhdunarodnoj konferencii po komp'yuternoj i kognitivnoj lingvistike TEL-2018 v 2-x tomah, Kazan. 2018. S. 393–405. EDN SQGRVY.

19. *Falileeva M., Kirillovich A., Shakirova L., Nevzorova O., Lipachev E., Dyupina A.* OntoMathEdu Educational Mathematical Ontology: Prerequisites, Educational Levels and Educational Projections // CEUR Workshop Proceedings. 2020. Vol. 2784. P. 346–351.

20. *Falileeva M., Kirillovich A., Hevzorova O.A., Shakirova L.R., Lipachev E.K., Dyupina A.E.* Educational projection systems, levels and prerequisites of mathematical ontology OntoMathEdu // Russian Digital Libraries. 2021. V. 24, No. 3. P. 505–530. <https://doi.org/10.26907/1562-5419-2021-24-3-505-530>

21. *Nevzorova O.A., Falileeva M.V., Kirillovich A.V., Shakirova L.R., Lipachev E.K., Dyupina A.E.* Modeling of didactic relationships in the OntoMathEdu educational mathematical ontology // CEUR Workshop Proceedings. 2021. V. 2910. P. 11–21.

22. *Falileeva M.V., Kirillovich A.V., Nevzorova O.A., Shakirova L.R., Lipachyov E.K., Dyupina A.E.* Razrabotka obrazovatel'ny`x proekcij matematicheskoy ontologii OntoMathEdu // Nauchny`j servis v seti Internet. 2020. № 22. S. 604–611. <https://doi.org/10.20948/abrau-2020-26>. EDN WBNWIC

23. *Kirillovich A., Falileeva M., Nevzorova O., Lipachev E., Dyupina A., Shakirova L.* Prerequisite Relationships of the OntoMathEdu Educational Mathematical Ontology // J.C. Figueroa-García et al. (Eds.). Proceedings of the 8th Workshop on Engineering Applications (WEA 2021). Communications in Computer and Information Science. 2021. Vol. 1431. P. 517–524. [https://doi.org/10.1007/978-3-030-86702-7\\_44](https://doi.org/10.1007/978-3-030-86702-7_44)

24. *Nevzorova O., Shakirova L., Falileeva M., Nikolaev K.* Towards Ontology-Based Software Services for Teaching School Mathematics // L. Gómez Chova et al. (Eds.). Proceedings of the 15th International Technology, Education and Development Conference (INTED 2021), March 8–9, 2021. P. 5288–5297. <https://doi.org/10.21125/inted.2021.1085>

25. *Nevzorova O.A., Lipachev E.K., Nikolaev K.S.* Semantic services of the digital ecosystem OntoMath for mathematical education // *Russian Digital Libraries*. 2023. Vol. 26. No. 4. P. 538–564. <https://doi.org/10.26907/1562-5419-2023-26-4-538-569>

26. *Nikolaev K.S., Kirillovich A.V., Nevzorova O.A.* A corpus-based approach to elementary geometry knowledge test generation. // *INTED2020 Proceedings*. 2020. P. 6342–6348. <https://doi.org/10.21125/inted.2020.1710>

27. *Nevzorova O.A., Nikolaev K.S.* Semantic Annotation of Mathematical Formulas in PDF-Documents // *Russian Digital Libraries*. 2022. Vol. 25. No. 6. P. 616–639. <https://doi.org/10.26907/1562-5419-2022-25-6-616-639>

28. *Nikolaev K.S.* Service for generating educational cards of mathematical concepts for a distance learning course in geometry // *Russian Digital Libraries*. 2023. Vol. 26. No. 3. P. 365–377. <https://doi.org/10.26907/1562-5419-2023-26-3-365-377>

29. *Nevzorova O.A., Kirillovich A.V., Nikolaev K.S., Galiaskarova K.R.* Towards a Semantically Annotated Corpus of Educational Mathematical Texts in Russian (short paper) // *International Conference on Data Analytics and Management in Data Intensive Domains*. 2019. P. 299–305. *CEUR Workshop Proceedings*. 2019. Vol. 2523. P. 299–305.

## СВЕДЕНИЯ ОБ АВТОРЕ



**НИКОЛАЕВ Константин Сергеевич** – к. т. н., младший научный сотрудник Лаборатории суперкомпьютерного моделирования Отделения суперкомпьютерных систем и параллельных вычислений ФГБУ «Национальный исследовательский центр «Курчатовский институт»». Основные направления научных исследований: обработка естественного языка, искусственный интеллект.

**Konstantin Sergeevich NIKOLAEV** – Cand. Sc. (Technology), Junior Researcher at the Supercomputer Modeling Laboratory of the Department of Supercomputer Systems and Parallel Computing of the Federal State Budgetary Institution "National Research Center "Kurchatov Institute"". Major fields of scientific research are Natural Language processing, artificial intelligence.

email: konnikolaeff@yandex.ru

ORCID: 0000-0003-3204-238X

*Материал поступил в редакцию 1 марта 2024 года*

*Переработанная версия поступила в редакцию 11 июня 2024 года*