

TeXML: XML-нотация для TeX

Паращенко Олег

8 июля 2004 г.

TeXML

Санкт-Петербург

xmlhack.ru

Аннотация

В работе описывается один из способов преобразования XML-документов в формат \LaTeX . Для этого предлагается использовать TeXML, промежуточное представление нотации TeX в виде XML.

1 Зачем нужен TeXML

TeXML[1] — это промежуточный формат для представления нотации TeX в виде XML. Он упрощает создание печатных версий документов, когда выполнены следующие условия:

- исходные документы представлены в виде XML,
- для создания печатной версии используется \LaTeX ,
- XSLT — лучший способ преобразования XML-данных.

Известны несколько проектов, в которых XML преобразуется в \LaTeX с помощью XSLT. Все они сталкивались с тем, что XSLT неэффективен, если результат трансформации — не XML.

Эту проблему можно обойти, используя промежуточный этап. Вначале XSLT эффективно преобразует XML в TeXML, а затем полученный TeXML сериализуется в TeX.

2 Краткое описание TeXML

Основные конструкции языка TeXML видны на примере:

```
<TeXML>
  <cmd name="documentclass">
    <opt>a4paper</opt>
    <parm>minimal</parm>
  </cmd>
  <env name="document">
```

```

    Это&nbsp;&mdash; пример.
  </env>
</TeXML>

```

Результат сериализации:

```

\documentclass[a4paper]{minimal}
\begin{document}
Это~\textemdash{} пример.
\end{document}

```

Главные задачи сериализатора:

- замена специальных символов (например, «<» на «\textless»),
- обработка unicode (например, вывод «д» или «\cyrchar\cyrd{}»),
- разбиение лигатур (например, вместо «--- » — «-{}-{}-»).

Помимо этих существенных преобразований, TeXML предоставляет также расширенные возможности.

- В язык встроена поддержка структурных конструкций L^AT_EX: окружений, групп, команд с параметрами.
- Благодаря автоматическому форматированию создаваемого кода, итоговые документы хорошо выглядят.

3 Преимущества TeXML

Простой код на TeXML

```
<group><cmd name="it" gr="0"/>\пример</group>
```

соответствует такому TeX-фрагменту:

```
{\it \textbackslash{}пример}
```

Если изучить историю проектов tbook[2], xsltml[3], dblatex[4], db2latex[5] и других, создающих TeX-код с помощью XSLT, то выяснится, что во всех них были подобные проблемы:

- не всегда экранируются специальные символы,
- иногда забывается пробел после команды и получается «\it*текст*» вместо «\it *текст*»,
- может теряться открывающая или закрывающая фигурная скобка,
- из-за объективных сложностей, системы ориентированы на западно-европейскую кодировку.

Эти и другие проблемы не возникают при использовании TeXML.

4 Другие версии T_EXML

При поиске решения для публикации XML через L^AT_EX выяснилось, что Douglas Lovell предложил[6] подход T_EXML ещё в 1999-м году, и была даже реализация на языке Java. Этот проект, T_EXMLatté[7], был заброшен несколько лет назад, но от него осталась описание, которое послужило основой для новой версии T_EXML.

Существует также T_EXMLapis[8], обработчик подмножества T_EXML на языке Perl, но его возможности ограничены.

Список литературы

- [1] TeXML: an XML vocabulary for TeX <http://getfo.org/texml/>
- [2] The tbook system for XML Authoring <http://tbookdtd.sourceforge.net/>
- [3] XSLT MathML Library <http://xsltml.sourceforge.net/>
- [4] DocBook to LaTeX/ConTeXt Publishing <http://dblatex.sourceforge.net/>
- [5] DB2LaTeX XSL Stylesheets <http://db2latex.sourceforge.net/>
- [6] Douglas Lovell, TeXML: Typesetting XML with TeX <http://www.tug.org/TUG99-web/pdf/lovell.pdf>
- [7] T_EXMLatté <http://www.alphaworks.ibm.com/tech/texml>
- [8] T_EXMLapis <http://www.blueweb.com/us/chouser/proj/texmlapis/>