

Русский язык в L^AT_EX 2_ε*

Аннотация

В декабре 1998 года в рамках проекта команды L^AT_EX3 реализована поддержка русского языка в соответствии со стандартом. Все необходимые файлы распространяются сейчас вместе с L^AT_EXом. Они включены в пакет средств Cyrillic Bundle. Кроме того, используется пакет russianb, входящий в систему Babel. Полезные ресурсы находятся также в пакете дополнительных средств поддержки русского языка T2.

Содержание

1	Кодировка входного файла	1
2	Кодировка шрифтов	2
2.1	Русский документ	3
2.2	Английский документ	3
2.3	Итоги	4
3	Математика	4
3.1	Кириллические математические алфавиты	4
3.2	Имена функций согласно российской традиции	4
4	Печать текста согласно российской традиции	4
4.1	Номер	4
4.2	Кавычки	5
4.3	Тире	5
4.4	Дефисы и переносы слов	5
5	Нумерация русскими буквами	5
6	Команды с предопределённым текстом	6
7	Проблемы	7

1 Кодировка входного файла

Если во входном файле встречаются 8-битные символы, то посредством пакета inputenc необходимо указать их кодировку:

```
\usepackage[имя кодировки]{inputenc}
```

*© 2002 Владимир Сюткин. Замечания приветствуются: syutkin@ns.kinetics.nsc.ru

В среде MS Windows, в которой используется кодовая страница 1251, *имя кодировки* задаётся как cp1251:

```
\usepackage[cp1251]{inputenc}
```

В MS-DOS со стандартной русской кодовой страницей 866 надо указать cp866, а в системах, основанных на UNIX с поддержкой русского языка согласно RFC 1489, *имя кодировки* задаётся как ko18-r. Полный список кодировок, которые поддерживает Cyrillic Bundle, можно найти в руководстве cyrguide.

Входной файл может содержать текст с разными кодировками символов. Текущая кодировка объявляется прямо в документе посредством декларации

```
\inputencoding{имя кодировки}
```

2 Кодировка шрифтов

По умолчанию Л^AT_EX использует 7-битную кодировку текстовых шрифтов OT1, в которой нет русских букв. Стандартными кодировками шрифтов с кириллическими буквами являются 8-битные кодировки T2A, T2B, T2C и X2. Они охватывают практически все основанные на кириллице языки. Для документов на русском языке обычно используется кодировка T2A, хотя буквы русского алфавита содержатся во всех кириллических кодировках. Если же вы пишете документ, скажем, на якутском языке, то вам придётся использовать кодировку T2B. Узнать, в какой кодировке определены буквы различных национальных алфавитов, можно в руководстве cyrguide.

Кодировки T2A, T2B и T2C включают в себя буквы латинского языка и поэтому подходят для смешанных текстов.

По стандарту кодировка текстовых шрифтов указывается посредством пакета fontenc:

```
\usepackage[кодировка шрифта]{fontenc}
```

Можно указать через запятую сразу несколько кодировок. Та, что указана последней, станет кодировкой по умолчанию.

Кодировку шрифта можно поменять прямо в документе посредством декларации¹

```
\fontencoding{кодировка шрифта}
```

`\fontencoding` только объявляет кодировку. Для того, чтобы Л^AT_EX стал использовать шрифты в объявленной кодировке, надо ещё применить декларацию

```
\selectfont
```

Между `\fontencoding` и `\selectfont` не должно быть текста (допускаются только пробелы и декларации выбора других атрибутов шрифтов). Например, текст в области действия деклараций

```
\fontencoding{U}\fontfamily{psy}\selectfont text
```

будет напечатан шрифтом Adobe Symbol, имеющим кодировку U и гарнитуру psy.

¹кодировка шрифта должна быть стандартной, иначе её надо сначала объявить посредством fontenc

2.1 Русский документ

Для русских документов² кодировку шрифтов можно установить с помощью пакета `russianb`. Он является частью системы многоязыковой поддержки `babel` и загружается обычно следующим образом:

```
\usepackage[russian]{babel}
```

`russianb` не только устанавливает кодировку текстовых шрифтов T2A, но и включает правила переноса русских слов³. Кроме того, пакет `russianb` переопределяет команды с предопределённым текстом типа `Contents` или `Figure`, и вводит новые команды для печати ряда символов согласно российской традиции.

Если русский документ содержит целые абзацы английского текста, а не отдельные слова, то перед ними надо ставить любую из двух эквивалентных команд

```
\English      \Eng
```

Они включают правила переноса слов английского языка, что даёт более аккуратную вёрстку абзацев с английским текстом. После окончания английского текста надо поставить любую из двух эквивалентных команд

```
\Russian      \Rus
```

Они восстановят правила переноса для русского языка.

2.2 Английский документ

Для документов на английском языке, содержащих небольшие включения текста на русском языке, можно ограничиться подключением пакета `fontenc` в виде

```
\usepackage[T2A]{fontenc}
```

`fontenc` установит кодировку текстовых шрифтов T2A и \LaTeX сверстаёт русский текст, но без переносов в словах. Если вас волнует качество вёрстки (без переносов в словах могут появиться длинные или разреженные строки), то места переноса в русских словах можно задать командой `\-` или в декларации `\hyphenation`. Но такая процедура приемлема для небольшого количества русского текста.

Если же документ на английском языке содержит большие фрагменты русского текста, то лучше воспользоваться пакетом `babel` с двумя опциями:

```
\usepackage[russian,english]{babel}
```

Поскольку опция `english` стоит последней, основным языком документа будет английский. Перед абзацами на русском языке надо ставить команду `\Russian` или `\Rus`, а после них — `\English` или `\Eng`. Для вставки коротких фраз на русском языке можно пользоваться командой

```
\textcyrillic{русский текст}
```

²Основным языком документа считается тот, на котором печатается предопределённый текст (см. [раздел 6](#)). В английском документе, скажем, подпись к таблице начинается со слова `Table`, а не со слова `Таблица`, как должно быть в русском документе.

³Образцы допустимых переносов в словах русского языка подключаются на стадии создания файла формата. Если в файл формата не включены правила переноса на русском языке, то \LaTeX будет верстать документы без переносов в словах.

2.3 Итоги

Итак, для русских документов преамбула входного файла, скажем, в MS Windows, должна содержать декларации

```
\usepackage[cp1251]{inputenc}
\usepackage[russian]{babel}
```

а для английских документов с русским текстом в самом общем случае:

```
\usepackage[cp1251]{inputenc}
\usepackage[russian,english]{babel}
```

3 Математика

3.1 Кириллические математические алфавиты

Пакет `russianb` объявляет кириллические математические алфавиты. Кириллические буквы в математической моде можно использовать после загрузки пакета `mathtext` из T2 *перед* загрузкой пакета `babel` с опцией `russian`⁴. По умолчанию буквы берутся из шрифта прямого начертания (в отличие от латинских букв). Если нужен иной шрифт, например, с курсивным начертанием, то надо воспользоваться кириллическими аналогами стандартных математических алфавитов:

```
\cyrmathrm \cyrmathbf \cyrmathsf \cyrmathit \cyrmathtt
```

Замечу, что практически всегда можно обойтись без кириллических математических алфавитов, поскольку русские буквы не принято использовать в качестве имён переменных, а включение в формулы русского текста, скажем, в качестве индексов можно выполнить командой `\text` из пакета `amstext`.

3.2 Имена функций согласно российской традиции

Пакет `russianb` вводит команды для печати имён гиперболических, тригонометрических и ряда других функций, принятых в русской литературе. Все команды собраны в следующей таблице (в скобках рядом с командами находится то, что они печатают):

<code>\arctg</code> (arctg)	<code>\ch</code> (ch)	<code>\ctg</code> (ctg)	<code>\sh</code> (sh)
<code>\arcctg</code> (arcctg)	<code>\cosec</code> (cosec)	<code>\cth</code> (cth)	<code>\tg</code> (tg)
<code>\th</code> (th)	<code>\Prob</code> (P)	<code>\nod</code> (н.о.д.)	<code>\nok</code> (н.о.к.)
<code>\Proj</code> (Пр)	<code>\Variance</code> (D)	<code>\NOD</code> (НОД)	<code>\NOK</code> (НОК)

Команды, которые печатают русские буквы (`\nod`, `\nok`, `\NOD`, `\NOK` и `\Proj`), работают только после подключения пакета `mathtext`.

4 Печать текста согласно российской традиции

4.1 Номер

Пакет `russianb` вводит команду `\No`, которая печатает номер в привычном для нас виде №

⁴`mathtext` перестает работать, если *после* него загрузить пакет `hyperref`. Поскольку пакет `hyperref` переопределяет многие команды, его рекомендуют загрузить последним. Поэтому сначала надо загрузить пакеты, чьи команды переопределяет `hyperref`, затем сам `hyperref` и только затем пакеты поддержки русского языка.

4.2 Кавычки

В английском языке приняты одинарные и двойные кавычки в виде ‘...’ и “...”. В России приняты французские («...») и немецкие („...“) кавычки (они называются «ёлочки» и «лапки», соответственно). «Лапки» обычно используются внутри «ёлочек», например, «... наш гордый „Варяг“ ...».

После подключения пакета `russianb`, французские левые и правые кавычки набираются как лигатуры `<<` и `>>` или командами⁵ `"<` и `">`, а немецкие левые и правые кавычки набираются как лигатуры `,,` и ```` или командами `"`` и `"'`.

Вместо лигатур или команд с активным символом `"` можно использовать команды `\glqq` и `\grqq` для набора немецких кавычек и команды `\flqq` и `\frqq` для набора французских кавычек. Они определены в пакете `babel`.

4.3 Тире

Пакет `russianb` вводит три команды для набора тире, которые печатаются согласно российской традиции.

Команда `"---` используется для печати тире в тексте. Оно несколько короче английского длинного тире. Кроме того, команда задаёт небольшую жёсткую отбивку от слова, стоящего перед тире. При этом, само тире не отрывается от слова. После тире следует такая же отбивка от текста, как и перед тире. При наборе текста между словом и командой, за которым она следует, должен стоять пробел.

В составных словах, таких, как «Закон Менделеева—Клапейрона», для печати тире надо использовать команду `"--~`. Она ставит более короткое, по сравнению с английским, тире и позволяет делать переносы во втором слове. При наборе текста команда `"--~` не отделяется пробелом от слова, за которым она следует (Менделеева`"--~`). Следующее за командой слово может быть отделено от неё пробелом или перенесено на другую строку.

Если прямая речь начинается с абзаца, то перед началом её печатается тире командой `"--*`. Она печатает русское тире и жёсткую отбивку нужной величины перед текстом.

Вместо команд, начинающихся с символа `"`, можно использовать аналогичные им по действию, но более длинные в наборе, команды `\cdash---`, `\cdash--~` и `\cdash--*`, соответственно.

4.4 Дефисы и переносы слов

Для печати дефиса в составных словах введены две команды. Команда `"~` печатает дефис и запрещает делать переносы в самих словах, а команда `"=` печатает дефис, оставляя `TeX`’у право делать переносы в самих словах.

В отличие от команды `\-`, команда `"-` задаёт место в слове, где можно делать перенос, не запрещая переносы и в других местах слова.

Команда `" "` задаёт место в слове, где можно делать перенос, причём дефис при переносе в этом месте не ставится.

Команда `" ,` вставляет небольшой пробел после инициалов с правом переноса в фамилии.

5 Нумерация русскими буквами

Пакет `russianb` определяет команды

⁵«Наши» команды начинаются с символа `"`, а не с `\` как принято в `TeX`.

`\asbuk{счётчик}` `\Asbuk{счётчик}`

для печати значения любого *счётчика* русскими строчными и прописными буквами, соответственно. Например, после переопределения

`\renewcommand{\theequation}{\asbuk{equation}}`

имеем

$$I = U + PV \quad (a)$$

$$\Psi = U - TS \quad (б)$$

Введены также *форматы*

`asbuk` `Asbuk`

для декларации

`\pagenumbering{формат}`

Они задают печать номера страницы русскими строчными и прописными буквами, соответственно.

6 Команды с предопределённым текстом

Пакет `russianb` переопределяет команду `\today` так, чтобы она печатала названия месяцев по русски. Например, 31 октября 2002 г.

Переопределены также команды из стандартных классов и некоторых пакетов, которые печатают предопределённый текст. Они приведены ниже вместе с английским и русским текстом:

<code>\abstractname</code>	Abstract	Аннотация
<code>\appendixname</code>	Appendix	Приложение
<code>\bibname</code>	Bibliography	Литература
<code>\chaptername</code>	Chapter	Глава
<code>\contentsname</code>	Contents	Содержание (в классе <code>article</code>) Оглавление (в классах <code>book</code> и <code>report</code>)
<code>\figurename</code>	Figure	Рис.
<code>\indexname</code>	Index	Предметный указатель
<code>\listfigurename</code>	List of Figures	Список иллюстраций
<code>\listtablename</code>	List of Tables	Список таблиц
<code>\partname</code>	Part	Часть
<code>\refname</code>	References	Список литературы
<code>\tablename</code>	Table	Таблица

Класс `letter`:

<code>\ccname</code>	cc	исх.	<code>\enclname</code>	encl	вкл.
<code>\pagename</code>	Page	с.	<code>\headtoname</code>	To	вх.

Пакет `makeidx`:

<code>\seename</code>	see	см.	<code>\alsiname</code>	see also	см. также
-----------------------	-----	-----	------------------------	----------	-----------

Пакет `amsthm`:

<code>\proofname</code>	Proof	Доказательство
-------------------------	-------	----------------

Определены также две новые команды

<code>\prefacename</code>	Предисловие	<code>\authorname</code>	Именной указатель
---------------------------	-------------	--------------------------	-------------------

Вы можете переопределить команды посредством декларации `\renewcommand`. Например, после переопределения в теле документа

```
\renewcommand{\indexname}{Алфавитный указатель}
```

перед указателем терминов будет напечатано «Алфавитный указатель», а не «Предметный указатель».

7 Проблемы

Читая входной файл, \LaTeX переводит русские буквы в специальные команды⁶. Именно в этом представлении пишется русский текст в служебные `AUX`, `TOC`, `LOF`, `LOT` и `IDX` файлы. В результате, «нормальный» пользователь не в состоянии редактировать их содержание в обычном текстовом редакторе. Естественно, такой `idx`-файл нельзя обрабатывать сразу программой `makeindex`, поскольку она будет сортировать термины на русском языке не по буквам, а по именам команд. Правильный `ind`-файл можно получить, обработав `idx`-файл специальным командным файлом, входящим в средства поддержки русского языка T2. Он сначала переводит русские буквы из представления \LaTeX в нормальное представление для текстовых файлов, а затем обрабатывает его программой `makeindex`.

Перейдём теперь к проблеме с русскоязычными документами, которая не решена до сих пор. В окно выполнения программы и в файл протокола (`log`-файл) \LaTeX пишет, используя свою кириллическую кодировку шрифтов. Она не соответствует ни одной из кириллических кодовых страниц. Поэтому, зачастую искать то место в документе, в котором \LaTeX выводит предупреждение или сообщение об ошибке, приходится по номеру строки, а не по тексту. Проще всего тем пользователям, которые работают в среде MS Windows, поскольку расположение русских букв в кодировке шрифтом T2A практически совпадает с их расположением в кодовой странице 1251 (буква ё не в счёт). Тем, кто работает в операционной системе с кодировкой символов согласно кодовым страницам 866 или `koï8-r`, приходится для просмотра `log`-файла использовать текстовый редактор, который переводит символы из кодировки 1251 в их кодировку.

⁶«Таблица» замещения загружается пакетом `inputenc` (в случае опции `cp1251` — из файла `cp1251.def`, входящего в `Cyrillic Bundle`).