

# **Введение в TEI Lite**

Лу Бернард (Lou Burnard),  
С.М. Шперберг - Мак-Куин (С.М. Sperberg-McQueen)

июнь 1995

## Содержание

<b>1 Введение</b> .....	<b>2</b>
<b>2 Пример</b> .....	<b>4</b>
<b>3 Структура текста TEI</b> .....	<b>7</b>
<b>4 Кодирование основной части</b> .....	<b>8</b>
4.1 Элементы для разделения текста .....	8
4.2 Заголовки и заключения .....	10
4.3 Проза, стихи и пьесы .....	10
<b>5 Номера строк и страниц</b> .....	<b>12</b>
<b>6 Как выделять фразы</b> .....	<b>13</b>
6.1 Изменение начертания шрифта и т.д. ....	13
6.2 Цитаты и связанные с ними особенности .....	14
6.3 Слова и выражения на иностранных языках .....	15
<b>7 Примечания</b> .....	<b>15</b>
<b>8 Перекрестные ссылки и связи</b> .....	<b>16</b>
8.1 Простые перекрестные ссылки .....	16
8.2 Расширенные указатели .....	18
8.3 Атрибуты связывания .....	20
<b>9 Редакторская правка</b> .....	<b>21</b>
<b>10 Пропуски, удаления и дополнения</b> .....	<b>21</b>
<b>11 Имена, даты, числа и сокращения</b> .....	<b>23</b>
11.1 Имена и называющие строки .....	23
11.2 Даты и время .....	24
11.3 Числа .....	25
11.4 Сокращения и их полные формы .....	25
11.5 Адреса .....	26
<b>12 Списки</b> .....	<b>26</b>
<b>13 Библиографические ссылки</b> .....	<b>28</b>
<b>14 Таблицы</b> .....	<b>29</b>
<b>15 Рисунки и графика</b> .....	<b>30</b>
<b>16 Интерпретация и анализ</b> .....	<b>31</b>
16.1 Орфография .....	31
16.2 Элементы интерпретации общего назначения .....	32
<b>17 Техническая документация</b> .....	<b>34</b>
17.1 Дополнительные элементы для технических документов .....	34
17.2 Генерируемые разделы .....	36
17.3 Генерирование указателя .....	36
<b>18 Национальные алфавиты, диакритические знаки и т.п.</b> .....	<b>37</b>
<b>19 Вводная и закрывающая части</b> .....	<b>39</b>

19.1 Вводная часть .....	39
19.1.1 Титульный лист .....	39
19.1.2 Вступительная часть .....	40
19.2 Закрывающая часть .....	41
19.2.1 Структурные подразделения закрывающей части .....	41
<b>20 Электронный титульный лист .....</b>	<b>42</b>
20.1 Описание файла .....	42
20.1.1 Сведения о названии .....	43
20.1.2 Сведения о редакциях текста .....	43
20.1.3 Сведения о размере .....	43
20.1.4 Сведения о публикации .....	44
20.1.5 Сведения о сериях и примечаниях .....	44
20.1.6 Описание источников .....	44
20.2 Описание кодирования .....	45
20.2.1 Описание проекта и способа отбора текстов .....	45
20.2.2 Определение способов редактирования .....	45
20.2.3 Объявление тегов, ссылок и классификаций .....	46
20.3 Описание параметров .....	47
20.4 Описание исправлений .....	48
<b>21 Список рассмотренных элементов .....</b>	<b>49</b>
21.1 Глобальные атрибуты .....	49
21.2 Элементы TEI Lite .....	49
<b>22 Библиография .....</b>	<b>54</b>

В этом документе приведены общие сведения об использовании системы кодирования текстов (TEI), дано описание управляемой пользователем части системы кодирования TEI. С помощью описанных здесь процедур можно получить разнообразные эффекты, обычно применяемые при оформлении текстов, причем так, чтобы электронные копии были максимально удобны и просты для использования, и чтобы обеспечить возможность использования этих копий в различных компьютерных системах. Описанная схема полностью совместима с полной системой TEI, как определено документом по TEI P3, *Guidelines for Electronic Text Encoding and Interchange* (Принципы кодирования и обмена электронными текстами), который был опубликован в Чикаго и Оксфорде в мае 1994 г. (Текущую версию этого документа можно найти через World Wide Web по адресу <http://www-tei.uic.edu/orgs/tei/intros/teiu5.tei> или <ftp://info.ox.ac.uk/pub/ota/TEI/doc/teiu5.tei>, а также на других сайтах - “зеркала” указанных. Этот документ доступен также в форме HTML по адресу <http://www-tei.uic.edu/orgs/tei/intros/teiu5.html> и <http://info.ox.ac.uk/archive/teilight/teiu5.html>, Образцы формальных определений типов документов SGML для описанного здесь набора тегов можно найти там же, в файле **teilight.dtd**: <ftp://www-tei.uic.edu/orgs/tei/p3/dtd/teilight.dtd> и <ftp://info.ox.ac.uk/pub/ota/TEI/dtd/teilight.dtd>.)

## **1 Введение**

Система кодирования текстов (TEI) направлена на обеспечение обмена информацией, хранимой в электронной форме. Основное внимание уделяется текстовой информации, но предусмотрены средства и для других форм, например, для графических изображений и звуковой информации. “Принципы” одинаково применимы при создании новых ресурсов и обмене существующими.

“Принципы” обеспечивают средства, позволяющие сделать ясными определенные особенности текста, причем так, чтобы облегчить обработку этого текста компьютерными программами, работающими на разных машинах. Процедуру выявления особенностей текста мы называем **разметкой** или **кодированием**. Любое представление текста на компьютере использует ту или иную форму разметки; одной из причин разработки системы TEI является существование огромного количества взаимно несовместимых систем кодирования, а также увеличение областей использования электронных текстов.

В “Принципах TEI” для определения схемы кодирования использован **стандартный язык обобщенной разметки** (SGML). SGML является международным стандартом (ISO 8879), все шире используемым в индустрии обработки информации. SGML позволяет формально определить схему кодирования в терминах **элементов** и **атрибутов**, а также с помощью правил, управляющих их размещением в тексте. Использование SGML в TEI усложняет систему, но делает ее универсальной, поскольку нет принципиальных отличий от других схем разметки SGML, и благодаря этому любое программное обеспечение общего назначения, которое может работать с SGML, может обрабатывать TEI-совместимые тексты.

TEI поддерживают Association for Computers and the Humanities (Ассоциация по компьютерам и гуманитарным наукам), Association for Computational Linguistics (Ассоциация по вычислительной лингвистике) и Association for Literary and Linguistic Computing (Ассоциация по компьютерным технологиям в литературе и лингвистике). Финансирование осуществляется U.S. National Endowment for the Humanities (Американский Национальный фонд пожертвований на гуманитарные науки), Directorate General XIII of the Commission of the European Communities (XIII Управление комиссии Европейского сообщества), Andrew W. Mellon Foundation (Фонд Эндрю У. Меллона) и Social Science and Humanities Research Council of Canada (Совет по социальным наукам и гуманитарным исследованиям Канады). Принципы системы TEI опубликованы в мае 1994 года, после шести лет разработки, в которую были вовлечены многие сотни ученых различных специальностей из разных стран мира.

В начале разработки общие цели системы TEI были сформулированы в закрытом отчете конференции по планированию работ, проведенной в Vassar College (Колледж Вассар), Нью-Йорк, в ноябре 1987 года; по этим “наброскам” был затем разработан ряд проектировочных документов. Согласно этим документам, в основу TEI должны быть положены следующие принципы:

- возможность получать в тексте эффекты, необходимые для исследовательской работы;
- простота, ясность и конкретность;
- несложность для использования без специализированного программного обеспечения;
- возможность точного определения и эффективной обработки текстов;
- возможность расширений, определяемых пользователем;
- соответствие существующим и новым стандартам.

Мир гуманитарных наук велик и разнообразен. Чтобы система с данными принципами стала широко распространенной, важно гарантировать, что:

1. общее ядро текстовых функций легко выделяется;
2. дополнительные специальные функции можно легко добавить в текст (или удалить из него);
3. имеется возможность нескольких параллельных кодировок одной и той же функции;
4. “богатство” разметки определяется пользователем, а набор минимальных необходимых условий очень невелик;
5. доступна соответствующая документация по текстам и их кодированию.

В настоящем документе, который называется “Введение в TEI Lite”, описано только небольшое подмножество из обширного набора элементов SGML и приведены рекомендации, вытекающие из целей этой разработки.

Была предпринята попытка выделить из нескольких сотен элементов SGML, определенных полной схемой TEI, полезный “стартовый набор”, в который вошли бы элементы, необходимые почти каждому пользователю. Опыт работы с TEI Lite будет чрезвычайно полезен для понимания полной системы TEI DTD и при определении того, какие необязательные части полной системы DTD необходимы для работы с конкретными типами текстов.

Цели, поставленные при выборе этого подмножества элементов, можно сформулировать следующим образом:

- это подмножество должно содержать большую часть набора основных тегов TEI, поскольку в нем содержатся элементы, соответствующие фактически всем типам текстов и всем видам процедур по обработке текстов;
- набор выбранных элементов должен обеспечивать адекватную обработку достаточно разнообразных текстов, с уровнем детальности, который задается существующей практикой (как демонстрируется, например, материалами Oxford Text Archive (Оксфордского Архива));
- он должен обеспечивать как создание новых документов, так и выполнение кодирования существующих;
- он должен быть таким, чтобы его можно было использовать с разнообразными существующими программными продуктами SGML;
- он должен получаться из полной системы TEI DTD с помощью механизмов расширения, которые описаны в “Принципах TEI”;

- он должен быть настолько мал и прост, насколько это совместимо с другими целями.

Читатель может сам судить о том, насколько удалось достичь этих целей. Во время написания уверенность разработчиков в том, что поставленные цели достигнуты хотя бы частично, подкреплялась практическим использованием результатов для кодирования реальных текстов. Оксфордский Архив использует TEI Lite, когда переводит тексты хранимых документов из их исходных разметок в SGML; Центры электронных текстов университета Вирджинии и Мичиганского университета используют TEI Lite для кодирования своих документов. И в самой системе TEI используется TEI Lite, в ее текущей технической документации — включая этот документ.

Хотя этот документ писался как автономный, что обычно для учебных текстов, читатель должен понимать, что здесь приведены далеко не все подробности схемы кодирования TEI. Все элементы, описанные здесь, полностью документированы в “Принципах TEI”, которые следует рассматривать как авторитетный источник не только по этим элементам, но и по многим другим, не упомянутым в настоящем документе. Предполагается, что читатель имеет некоторые базисные знания SGML.

## 2 Пример

Приведем краткий пример, с помощью которого покажем, что получается, когда отрывок текста введен в компьютер без учета всех требований разметки или возможностей электронных текстов. В идеале такой текст можно создать с помощью очень точного оптического сканера. Полученный текст будет точно соответствовать печатному оригиналу, в нем сохранятся такие же строки, как в исходном тексте, будут вставлены пробелы, чтобы сохранить исходную компоновку заголовков, исходные страницы и так далее. Если были нужны символы, отсутствующие на клавиатуре (например, буква **a** с ударением в слове **faál**, или длинное тире), предпринималась попытка имитировать их.

### CHAPTER 38

READER, I married him. A quiet wedding we had: he and I, the parson and clerk, were alone present. When we got back from church, I went into the kitchen of the manor-house, where Mary was cooking the dinner, and John cleaning the knives, and I said --

'Mary, I have been married to Mr Rochester this morning.' The housekeeper and her husband were of that decent, phlegmatic order of people, to whom one may at any time safely communicate a remarkable piece of news without incurring the danger of having one's ears pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy wonderment. Mary did look up, and she did stare at me; the ladle with which she was basting a pair of chickens roasting at the fire, did for some three minutes hang suspended in air, and for the same space of time John's knives also had rest from the polishing process; but Mary, bending again over the roast, said only --

'Have you, miss? Well, for sure!'

A short time after she pursued, 'I seed you go out with the master, but I didn't know you were gone to church to be wed'; and she basted away. John, when I turned to him, was grinning from ear to ear.

'I telled Mary how it would be,' he said: 'I knew what Mr Edward' (John was an old servant, and had known his master when he was the cadet of the house, therefore he often gave him his Christian name) -- 'I knew what Mr Edward would do; and I was certain he would not wait long either: and he's done right, for aught I know. I wish you joy, miss!' and he politely pulled his forelock.

'Thank you, John. Mr Rochester told me to give you and Mary this.'

I put into his hand a five-pound note. Without waiting to hear more, I left the kitchen. In passing the door of that sanctum some time after, I caught the words --

'She'll happen do better for him nor ony o' f grand ladies.' And again, 'If she ben't one o' th' handsomest, she's noan faa''l, and vary

good-natured; and i' his een she's fair beautiful, onybody may see that.'

I wrote to Moor House and to Cambridge immediately, to say what I had done: fully explaining also why I had thus acted. Diana and

474

JANE EYRE

475

Mary approved the step unreservedly. Diana announced that she would just give me time to get over the honeymoon, and then she would come and see me.

'She had better not wait till then, Jane,' said Mr Rochester, when I read her letter to him; 'if she does, she will be too late, for our honeymoon will shine our life long: its beams will only fade over your grave or mine.'

How St John received the news I don't know; he never answered the letter in which I communicated it: yet six months after he wrote to me, without, however, mentioning Mr Rochester's name or alluding to my marriage. His letter was then calm, and though very serious, kind. He has maintained a regular, though not very frequent correspondence ever since: he hopes I am happy, and trusts I am not of those who live without God in the world, and only mind earthly things.

В этой записи имеется ряд недостатков:

- номера страниц и колонтитулы идут прямо в тексте, из-за этого программам обработки трудно различать их;
- нет разницы между одинарными кавычками и апострофами, поэтому трудно точно определить, какие фразы являются прямой речью;
- сохранение переносов, которые были в исходном тексте, означает, что простые программы поиска не смогут обнаружить слово, если оно оказалось разорванным;
- буква с ударением в слове **faál** и длинное тире представлены специальными группами символов, которые не соответствуют стандарту, и поэтому корректно обрабатываются только если устройство воспроизведения данных специально настроено;
- абзацы выделены только отступом, а знак возврата каретки, жестко задающий формат, имеется в конце каждой строки. Следовательно, если изменить размер шрифта, которым печатается текст, то перереформатирование текста будет непростой задачей.

Теперь покажем, как будет выглядеть тот же самый отрывок, если использовать кодирование в соответствии с принципами TEI. Как будет видно далее, существует много способов расширить эти правила кодирования, но как минимум, подход с использованием TEI позволяет добиться следующих результатов:

- Разделение на абзацы теперь выполнено явно.
- Апострофы отличаются от кавычек.
- Для буквы с ударением и длинного тире использованы объектные ссылки.
- Концы страниц отмечены одним пустым элементом **<pb>**.
- Чтобы упростить поиск и обработку, не сохранены такие же строки, как в оригинале, а слова, разорванные в конце строк, перенесены целиком на следующую строку без комментариев. Если бы было необходимо сохранить строки оригинала, это было бы несложно сделать, хотя в этом примере строки не сохранены.
- Для удобства читки корректуры перед каждым абзацем добавлена пустая строка, а отступ удален.

```
<pb n='474'>
```

```
<div1 type=chapter n='38'>
```

```
<p>Reader, I married him. A quiet wedding we had: he and I, the parson and clerk, were alone present. When we got back from church, I went into the kitchen of the manor-house, where Mary was cooking the dinner, and John cleaning the knives, and I said &dash;
```

<p><q>Mary, I have been married to Mr Rochester this morning. </q> The housekeeper and her husband were of that decent, phlegmatic order of people, to whom one may at any time safely communicate a remarkable piece of news without incurring the danger of having one's ears pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy wonderment. Mary did look up, and she did stare at me; the ladle with which she was basting a pair of chickens roasting at the fire, did for some three minutes hang suspended in air, and for the same space of time John's knives also had rest from the polishing process; but Mary, bending again over the roast, said only &dash;

<p><q>Have you, miss? Well, for sure!</q>

<p>A short time after she pursued, <q>I seed you go out with the master, but I didn't know you were gone to church to be wed</q>; and she basted away. John, when I turned to him, was grinning from ear to ear. <q>I telled Mary how it would be,</q> he said: <q>I knew what Mr Edward</q> (John was an old servant, and had known his master when he was the cadet of the house, therefore he often gave him his Christian name) &dash; <q>I knew what Mr Edward would do; and I was certain he would not wait long either: and he's done right, for aught I know. I wish you joy, miss!</q> and he politely pulled his forelock.

<p><q>Thank you, John. Mr Rochester told me to give you and Mary this.</q>

<p>I put into his hand a five-pound note. Without waiting to hear more, I left the kitchen. In passing the door of that sanctum some time after, I caught the words &dash;

<p><q>She'll happen do better for him nor ony o' t' grand ladies.</q> And again, <q>If she ben't one o' th' handsomest, she's noan fa&agrave;l, and varry good-natured; and i' his een she's fair beautiful, onybody may see that.</q>

<p>I wrote to Moor House and to Cambridge immediately, to say what I had done: fully explaining also why I had thus acted. Diana and <pb n='475'> Mary approved the step unreservedly. Diana announced that she would just give me time to get over the honeymoon, and then she would come and see me.

<p><q>She had better not wait till then, Jane,</q> said Mr Rochester, when I read her letter to him; <q>if she does, she will be too late, for our honeymoon will shine our life long: its beams will only fade over your grave or mine.</q>

<p>How St John received the news I don't know: he never answered the letter in which I communicated it: yet six months after he wrote to me, without, however, mentioning Mr Rochester's name or alluding to my marriage. His letter was then calm, and though very serious, kind. He has maintained a regular, though not very frequent correspondence ever since: he hopes I am happy, and trusts I am not of those who live without God in the world, and only mind earthly things.

Возможность концентрации внимания на самом тексте Бронте, а не на особенностях его печати в данном конкретном издании - это один из аспектов основной цели кодирования: ее избирательность.

Кодирование выявляет только те особенности текста, которые интересуют кодирующего. Даже на этом коротком примере легко представить себе, как можно улучшить кодирование. Например:

- можно привести правильную форму диалектных выражений;
- можно добавить к любому отрывку сноски с толкованиями или комментариями
- можно добавить указатели, связывающие части этого текста с другими;
- различные типы имен собственных можно выделить из окружающего текста;
- подробная библиографическая информация о происхождении и связанных с текстом обстоятельствах может быть предпослана ему;
- можно обеспечить лингвистический анализ отрывка по отдельным предложениям, по частям сложных предложений, по словам и так далее, причем каждая единица связывается с соответствующими кодами категорий;
- текст можно разделить на сюжетные или логические единицы;
- в систему кодирования могут быть включены систематический анализ или интерпретация текста, с возможной сложной выверкой или связью между текстом и анализом, либо между текстом и одним или несколькими его переводами;
- отрывки текста можно связать с изображениями или звуковыми ресурсами, которые хранятся на других носителях.

Рекомендации по выполнению всего этого с помощью системы ТЕИ описаны далее в этом документе. В полной схеме ТЕИ предусмотрено множество других возможностей, среди которых, например, такие:

- подробный анализ компонентов имен;
- подробная мета-информация, обеспечивающая информацию о тезаурусе и стиле по первоисточникам или темам текста;
- информация об истории печати или вариантах рукописей, представленная определенным рядом версий текста.

За подробными сведениями об этих и многих других возможностях следует обратиться к полному варианту “Принципов”.

### 3 Структура текста ТЕИ

Все тексты в формате ТЕИ содержат (а) **заголовок ТЕИ** (размечаемый как элемент `<teiHeader>`) и (б) собственно текст (размечаемый как элемент `<text>`).

В заголовке ТЕИ содержится информация, аналогичная той, что находится на титульном листе печатного текста. Она может содержать до четырех частей: библиографическое описание машиночитаемого текста, описание способа кодирования, неблиографическое описание текста (**параметры текста**), и архив исправлений. Более подробные сведения о заголовке приведены в разделе 20.

Текст ТЕИ может быть **монолитным** (отдельное произведение) или **объединенным** (набор отдельных произведений, как, например, в антологии). В любом случае текст может иметь необязательные **вводную часть** и **закрывающую часть**. Между ними располагается **основная часть** текста, которая, в случае объединенного текста, может состоять из **групп**, а они, в свою очередь, из других групп или текстов.

В результате кодирования монолитного текста получается структура, подобная приведенной ниже:

```
<TEI.2>
  <teiHeader> [ Информация заголовка ТЕИ ] </teiHeader>
  <text>
    <front> [ вводная часть ... ] </front>
    <body> [ тело текста ... ] </body>
    <back> [ закрывающая часть ... ] </back>
  </text>
</TEI.2>
```

Объединенный текст также имеет необязательные вводную и закрывающую части. Между ними располагается одна или несколько групп текстов, каждая из которых тоже может иметь свои собственные вводные и завершающие части. Таким образом, в результате кодирования объединенного текста получается структура, подобная приведенной ниже:

```
<TEI.2>
  <teiHeader> [ информация заголовка объединенного текста ] </teiHeader>
  <text>
    <front> [ вводная часть объединенного текста ] </front>
    <group>
```



```

    <text>
      <front> [ вводная часть первого текста ] </front>
      <body> [ тело первого текста ] </body>
      <back> [ заключительная часть первого текста ] </back>
    </text>
    <text>
      <front> [ вводная часть второго текста ] </front>
      <body> [ тело второго текста ] </body>
      <back> [ заключительная часть второго текста ] </back>
    </text>
    [ другие группы или тексты ]
  </group>
  <back> [ заключительная часть объединенного текста ] </back>
</text>
</TEI.2>

```

Можно также определить композицию текстов TEI, каждый из них со своим собственным заголовком. Такой набор называется **совокупностью текстов TEI** и может иметь свой заголовок:

```

<teiCorpus>
  <teiHeader> [заголовок совокупности] </teiHeader>
  <TEI.2>
    <teiHeader> [заголовок первого текста] </teiHeader>
    <text> [первый текст совокупности] </text>
  </TEI.2>
  <TEI.2>
    <teiHeader> [заголовок второго текста] </teiHeader>
    <text> [второй текст совокупности] </text>
  </TEI.2>
</teiCorpus>

```

Однако, нельзя создать композицию из совокупностей, то есть набор из нескольких элементов **<teiCorpus>**, рассматриваемый как единый объект. Это ограничение действует в текущей версии “Принципов TEI”.

В остальной части настоящего документа рассматриваются в основном простые текстовые структуры. При обсуждении каждой структуры приводится краткий список соответствующих **элементов TEI** и их описания, а затем определения всех **атрибутов**, характерных для данного элемента. В большинстве случаев приводятся примеры.

#### 4 Кодирование основной части

Как было указано выше, простой документ TEI на текстовом уровне состоит из следующих элементов:

- <front>** содержит различную вступительную информацию (заголовки, титульный лист, предисловия, посвящения и т.п.), которую размещают перед основным текстом.
- <group>** содержит число монолитных текстов или групп текстов.
- <body>** содержит всю основную часть одного монолитного текста, исключая то, что относится к вводной или закрывающей частям текста.
- <back>** содержит различные приложения и т.п., которые располагаются после основной части текста.

Элементы, характерные для вводной и завершающей частей описаны ниже, в разделе 19. В данном разделе обсуждаются элементы, размечающие основную часть текста.

##### 4.1 Элементы для разделения текста

Основная часть текста (написанного прозой) может представлять собой либо просто набор абзацев, либо эти абзацы могут быть сгруппированы в главы, разделы, подразделы и т.д. В первом случае каждый абзац отмечается тегом **<p>**. Во втором случае элемент **<body>** можно разделить либо на ряд элементов **<div1>**, либо на ряд элементов **<div>**, каждый из которых тоже можно разделить на более мелкие единицы, как описано ниже:

- <p>**       отмечает абзацы написанного прозой текста.
- <div>**       содержит раздел вводного, основного или закрывающего текста.
- <div1>**       содержит раздел первого уровня вводного, основного или закрывающего текста (это самый высокий уровень, если элемент **<div0>** не используется, или второй уровень, если он используется).

Если необходимы структурные разделы более мелкие, чем **<div1>**, то элемент **<div1>** можно разделить на элементы **<div2>**, а элементы **<div2>** на еще меньшие элементы **<div3>** и т.д., до уровня **<div7>**. Если в структуре имеется более семи уровней, следует либо изменить набор тегов TEI, чтобы воспринимались элементы **<div8>** и т.д., либо использовать нумерованный элемент **<div>**, поскольку его можно разделять на более мелкие элементы **<div>** без ограничений на количество уровней.

Для все этих элементов **разделов** необходимы следующие три атрибута:

- type**   Этот атрибут дает принятое название для данной категории раздела текста. Типичные значения этого атрибута - "Book" (книга), "Chapter" (глава), "Poem" (стихотворение) и т.д. Среди других возможных значений - "Group" (группа) для групп стихотворений и т.п., рассматриваемых как одна единица, "Sonnet" (сонет) "Speech" (речь) и "Song" (песня). Следует отметить, что любое значение атрибута **type**, присвоенное первому из элементов **<div>**, **<div1>**, **<div2>** и т.д. в тексте, предполагается относящимся ко всем последующим элементам **<div>**, **<div1>** и т.д. в пределах одного и того же элемента **<body>**. Это означает, что следует задавать значение для первого элемента раздела каждого типа, или при каждом изменении значения.
- id**      Этот атрибут представляет собой уникальный идентификатор раздела, который можно использовать для перекрестных ссылок или других связей с этим разделом, например, для комментариев, как это описано в разделе 8. Часто бывает полезным использовать атрибут **id** для каждой большой структурной единицы в тексте и присваивать значения идентификатора по какой-либо системе, например, дописывать номер раздела к короткому коду из названия рассматриваемого произведения, как это сделано в приведенных ниже примерах.
- n**       Атрибут **n** означает краткое (но понятное) название или номер раздела, которые можно использовать вместо идентификатора для обозначения этого раздела. Если уже существует общепринятая форма ссылок или сокращений для частей некоторого текста (например, шаблон книга/глава/стих для цитат из Библии), то именно в атрибуте **n** можно зафиксировать это.

Атрибуты **id** и **n** настолько полезны на практике, что они предусмотрены для любого элемента в TEI DTD: они являются **глобальными атрибутами**. Другие глобальные атрибуты, определенные в системе TEI Lite, рассмотрены в разделе 8.3.

Значение каждого атрибута **id** должно быть уникальным в пределах одного документа. Один из простых способов обеспечить это заключается в том, что атрибуту **id** присваиваются значения, отражающие иерархическую структуру документа. Например, *Богатство государств* Смита в первом издании состояло из пяти книг, каждая из которых разделялась на главы, а некоторые из глав были дополнительно разбиты на разделы. Для такой структуры можно определить значения атрибута **id** следующим образом:

```
<div1 id=WN1 n='I' type='book'>
  <div2 id=WN101 n='I.1' type='chapter'>
    ... </div2>
  <div2 id=WN102 n='I.2' type='chapter'>
    ... </div2>
    ...
  <div2 id=WN110 n='I.10' type='chapter'>
    <div3 id=WN1101 n='I.10.1' type='part'>
      ... </div3>
    <div3 id=WN1102 n='I.10.2' type='part'>
      ... </div3>
    </div2>
    ...
</div1>
<div1 id=WN2 n='II' type='book'>
  ....
```

```
</div1>
```

```
...
```

Для атрибутов **id** и **n** можно использовать и другую схему нумерации; это часто полезно там, где использована каноническая схема ссылок, которая не соответствует структуре произведения. Например, пусть роман разделен на книги, каждая из которых состоит из глав, причем главы пронумерованы последовательно во всем произведении, а не в пределах каждой книги. В этом случае можно воспользоваться следующей схемой:

```
<div1 id=TS01 n='1' type='Volume'>
  <div2 id=TS011 n='1' type='Chapter'>
    ...
  <div2 id=TS012 n='2'>
    ...
</div1>
<div1 id=TS02 n='2' type='Volume'>
  <div2 id=TS021 n='3' type='Chapter'>
    ...
  <div2 id=TS022 n='4'>
    ...
</div1>
```

В этом примере произведение состоит из двух томов, каждый из которых содержит две главы. Главы пронумерованы как обычно принято, от 1 до 4, но структура указанных значений **id** такова, что можно получить дополнительную информацию о месте главы в произведении, как если бы главы были пронумерованы 1.1, 1.2, 2.1, 2.2.

## 4.2 Заголовки и заключения

Каждый элемент **<div>**, **<div1>**, **<div2>** и т.д. может иметь название или заголовок вначале и (менее часто) заключение, например, “Конец главы 1”. Чтобы определить заголовки и заключения, можно воспользоваться следующими элементами:

- <head>** содержит любые заголовки, например, название раздела, заголовок списка или глоссария.
- <trailer>** содержит заключающее название или “окончание”, которое располагается после текста раздела.

Некоторые другие элементы, которые могут потребоваться в начале или в конце раздела текста, описаны ниже, в разделе 19.1.2.

Включать ли заголовки и заключения в транскрипцию - самостоятельно решает специалист, кодирующий текст. Там, где заголовки постоянные (например, “Глава 1”) или могут быть выражены как значение атрибута (например, **<div1 type='Chapter' n=1>**), их можно опустить; если же заголовки, напротив, содержат уникальный текст, их обязательно следует включить. Например, начало романа Харди *Под деревом зеленым* можно оформить следующим образом:

```
<div1 id=UGT1 n='Winter' type='Part'>
<div2 id=UGT11 n='1' type='Chapter'>
<head>Mellstock-Lane</head>
<p>To dwellers in a wood almost every species of tree ...
```

## 4.3 Проза, стихи и пьесы

Как отмечено выше, абзацы, составляющие отдельные части текста следует отмечать тегом **<p>**. Например:

```
<body>
<p>I fully appreciate Gen. Pope's splendid achievements
with their invaluable results; but you must know that
Major Generalships in the Regular Army, are not as
plenty as blackberries.
</p>
</body>
```

Предусмотрен ряд различных тегов для кодирования структурных компонентов стихов или текста ролей (пьесы, сценарии фильмов и т.п.):

- <l>** содержит одну (возможно, неполную) стихотворную строку. Включает следующие атрибуты:  
**part** указывает, является ли строка метрически полной. Допустимы следующие значения: **F** - для завершающей части неполной строки; **Y** - для метрически неполной строки; **N** - для полной строки, или если неважно, является ли строка полной или нет; **I** - для начальной части неполной строки; **M** - для средней части неполной строки.
- <lg>** содержит группу стихотворных строк, составляющих формальную единицу, например, строфу, рефрен, стихотворный абзац и т.д.
- <sp>** содержит прямую речь персонажа в тексте пьесы (сценария) или отрывок, представляющий таковую в тексте прозаического или стихотворного произведения. Включает следующие атрибуты:  
**who** указывает (с помощью идентификатора), кто произносит данный текст.
- <speaker>** содержит специальную форму заголовка или метки, где указано имя одного или нескольких персонажей, произносящих реплики в пьесе или в некотором фрагменте текста.
- <stage>** содержит различные ремарки к тексту пьесы или к отрывку пьесы. Включает следующие атрибуты:  
**type** определяет тип указаний для постановки (ремарок). Предлагаются следующие значения: **entrance** (о выходе на сцену), **exit** (об уходе со сцены), **setting** (о декорациях и костюмах), **delivery** (о манере декламации) и т.д.

Вот, например, начало стихотворения, в котором тегами отмечены стихотворные строки и строфы:

```
<lg n=I>
<l>I Sing the progresse of a
  deathlesse soule,</l>
<l>Whom Fate, with God made,
  but doth not controule,</l>
<l>Plac'd in most shapes; all times
  before the law</l>
<l>Yoak'd us, and when, and since,
  in this I sing.</l>
<l>And the great world to his aged evening; </l>
<l>From infant morne, through manly noone I draw.</l>
<l>What the gold Chaldee, of silver Persian saw,</l>
<l>Greeke brass, or Roman iron, is in this one;</l>
<l>A worke t'out weare Seths pillars, bricke and stone,</l>
<l>And (holy writs excepted) made to yeeld to none,</l>
</lg>
```

Следует отметить, что элемент **<l>** отмечает стихотворные, а не типографские строки; из-за этого исходное разбиение на строки в приведенных выше нескольких первых строчках не является очевидным при данном кодировании и может быть потеряно. Если необходимо разметить типографские строки, можно воспользоваться элементом **<lb>**, который описан в разделе 5.

Иногда, особенно в пьесах, стихотворные строки разделены на реплики нескольких персонажей. Самый простой способ кодирования в таких случаях - использование атрибута **part**; это позволяет определить, что строки, разбитые на указанные фрагменты, являются неполными, как в следующем примере:

```
<div1 type = 'Act' n = 'I'><head>ACT I</head>
<div2 type = 'Scene' n = '1'><head>SCENE I</head>
<stage rend=italic>
Enter Barnardo and Francisco, two Sentinels, at several doors</stage>
<sp><speaker>Barn<l part=Y>Who's there?
<sp><speaker>Fran<l>Nay, answer me. Stand and unfold yourself.
<sp><speaker>Barn<l part=i>Long live the King!
<sp><speaker>Fran<I part=m>Barnardo?
```

```
<sp><speaker>Barn<l part=f>He.
<sp><speaker>Fran<l>You come most carefully upon your hour.
```

Тем же способом можно воспользоваться, если на реплики двух персонажей разделены строфы:

```
<sp><speaker>First voice</speaker>
<lg type=stanza part=I>
<l>But why drives on that ship so fast
<l>Withouten wave or wind?
</lg>
<sp><speaker>Second Voice</speaker>
<lg part=F>
<l>The air is cut away before.
<l>And closes from behind.
</lg>
```

Этот пример показывает, как представить диалог из прозаического произведения в виде отрывка из пьесы. Из этого примера также видно, как использовать атрибут **who** для того, чтобы указать, кто произносит данные реплики в диалоге.

```
<sp who=OPI><speaker>The reverend Doctor Opimiam</speaker>
<p>I do not think I have named a single unrepresentable fish.
<sp who=GRM><speaker>Mr Gryll</speaker>
<p>Bream, Doctor: there is not much to be said for bream.
<sp who=OPI><speaker>The Reverend Doctor Opimiam</speaker>
<p>On the contrary, sir, I think there is much to be said for him.
In the first place ....
<p>Fish, Miss Gryll -- I could discourse to you on fish by
the hour: but for the present I will forbear...
</sp>
```

## 5 Номера строк и страниц

Конец строки и страницы можно отметить следующими пустыми элементами.

- <pb>** отмечает границу между двумя страницами текста в стандартной системе разметки.
- <lb>** отмечает начало новой (типографской) строки в определенном издании или редакции текста.

Эти элементы отмечают отдельное место в тексте, а не отрывок текста. Чтобы вставить номер страницы или строки, в теге следует воспользоваться глобальным атрибутом **n**. Кроме того, эти два элемента могут иметь следующий атрибут:

**ed** указывает издание или редакцию, в которых конец страницы находится в этом месте.

При работе с разбитым на страницы оригиналом часто бывает полезно зафиксировать это разбиение, хотя бы только для того, чтобы облегчить чтение корректуры. Фиксирование концов строк полезно по той же причине; однако, если в печатном оригинале имеются переносы в конце строк, необходимо отдельно рассмотреть вопрос о том, следует ли фиксировать такие строки.

Если имеется несколько изданий, каждое со своей нумерацией страниц и т.д., укажите издание, о котором идет речь, с помощью атрибута **ed** и вставьте необходимое количество тегов. Например, в приведенном ниже отрывке указано, где находится конец страницы в двух различных изданиях (**ED1** и **ED2**):

```
<p>I wrote to Moor House and to Cambridge immediately, to
say what I had done: fully explaining also why I had thus
acted. Diana and <pb ed=ED1 n='475'> Mary approved the
step unreservedly. Diana announced that she would
<pb ed=ED2 n='485'>just give me time to get over the
honeymoon, and then she would come and see me.
```

Элементы **<pb>** и **<lb>** являются частными случаями общего класса элементов **milestone** (закладка), которые отмечают “ориентиры” в тексте. В TEI Lite имеется общий элемент **<milestone>**, который не ограничен какими-либо специальными случаями, но может отмечать любой вид ориентира; например, конец

колонки, начало нового типа раздела, не отмеченного другими способами и т.д. Этот элемент имеет следующее описание и атрибуты:

**<milestone>** отмечает границы между разделами текста в соответствии со стандартной системой разметки. Включает следующие атрибуты:

**ed** указывает издание или редакцию, к которой относится закладка.

**unit** указывает, какой тип раздела изменяется на этой закладке.

Названия (имена) типов единиц текста и изданий, которые используются в атрибутах **ed** и **unit**, можно выбирать любые, но эти имена необходимо документировать в заголовке.

Элемент **<milestone>** можно использовать вместо других (“специализированных”) элементов, либо можно использовать набор специализированных элементов; однако, не следует произвольно смешивать элементы.

## 6 Как выделять фразы

### 6.1 Изменение начертания шрифта и т.д.

Выделенными словами или фразами называются те, которые визуально отличаются от остального текста (чаще всего другим типом шрифта, почерком, цветом чернил) с целью привлечь внимание читателя.

Глобальный атрибут **rend** может быть приписан любому элементу и использоваться там, где необходимо задать подробности выделения, используемого для данного элемента. Например, заголовок, выделенный полужирным шрифтом, можно отметить следующим образом: **head rend='Bold'**, а заголовок, выделенный курсивом, обозначить как **head rend='Italic'**.

Не всегда возможно указать какие-либо структурные признаки текста, из-за которых он подлежит выделению (как это сделано в приведенном выше примере, где выделен заголовок). В подобных случаях используется элемент **<hi>**, с помощью которого можно указать подлежащий выделению фрагмент текста, не определяя при этом статус выделяемого фрагмента (т.е. не указывая причину выделения текста).

**<hi>** отмечает слово или фразу, которые графически отличаются от остального текста, причем причина выделения не указывается.

В приведенном ниже примере шрифтами выделены подзаголовок и встречающееся в тексте имя собственное, причем эти изменения зафиксированы, но не объясняются:

```
<hi rend=gothic>And this Indenture further witnesseth</hi>  
that the said <hi rend=italic>Walter Shandy</hi>, merchant,  
in consideration of the said intended marriage ...
```

С другой стороны, для случаев, когда причина выделения ясна, предусмотрен ряд других, более специализированных элементов.

**<emph>** отмечает слова или фразы, выделяемые с целью получения лингвистического или риторического эффекта.

**<foreign>** указывает слово или фразу, написанные не на том же языке, что весь остальной текст.

**<mentioned>** указывает слова или фразы, упомянутые, но не использованные.

**<term>** содержит обозначение из слов (одного или нескольких) или символов, которое рассматривается как технический термин.

**<title>** содержит название произведения, т.е. статьи, книги, журнала или серии, включая все альтернативные названия или подзаголовки. Включает следующие атрибуты:

**level** указывает, является ли данный заголовок названием статьи, книги, журнала, серии или неопубликованного материала. Допустимы следующие значения: **m** - для монографических названий (книга, собрание сочинений или другое, опубликованное как самостоятельная единица, включая отдельные тома многотомных произведений); **s** (название серии); **j** (название журнала); **u** - для названий неопубликованных материалов (включая тезисы и диссертации, кроме опубликованных в коммерческой прессе); **a** - для аналитических названий (статья, стихотворение или другая единица, опубликованная как часть более крупной единицы).

**type** классифицирует названия в соответствии с принятой типологией.

Включены значения: **abbreviated**, **main**, **subordinate** (соответственно “сокращенное”, “главное”, “подчиненное”; предназначены для подзаголовков и названий частей), и **parallel** (“параллельное”, для альтернативных названий, часто на другом языке, под которыми также известно данное произведение).

Некоторые из характерные особенности (особенно цитаты и глоссы) можно найти в тексте отмеченными либо выделением, либо кавычками. В любом случае необходимы элементы `<q>` и `<gloss>` (их использование обсуждается в следующем разделе). Если выделение должно быть зафиксировано, следует воспользоваться глобальным атрибутом **rend**.

В качестве примера использования описанных выше элементов рассмотрим следующее предложение: On the one hand the *Nibelungenlied* is associated with the new rise of romance of twelfth-century France, the *romans d'antiquite*, the romances of Chretien de Troyes, and the German adaptations of these works by Heinrich van Veldeke, Hartmann von Aue, and Wolfram von Eschenbach (С одной стороны, *Песнь о Нибелунгах* связана с новым подъемом во Франции двенадцатого века рыцарского романа, *romans d'antiquite*, романами Кретьена де Тройе и немецкими переложениями этих произведений Генриха ван Вельдека, Гартмана фон Эйе и Вольфрама фон Эшенбаха). Если указывать причину выделения слов, предложение будет выглядеть следующим образом:

```
On the one hand the <title>Nibelungenlied</title> is associated
with the new rise of romance of twelfth-century France, the
<foreign>romans d'antiquite</foreign>, the romances of Chr&eacute;tien de Troyes,
...
```

Если же необходимо просто сохранить “внешность” оригинала, этот пример может выглядеть так:

```
On the one hand the <hi rend=italic>Nibelungenlied</hi>
is associated with the new rise of romance of twelfth-century
France, the <hi rend=italic>romans
d'antiquite</hi>, the romances of
Chr&eacute;tien de Troyes, ...
```

## 6.2 Цитаты и связанные с ними особенности

Кавычки, как и изменение шрифта, принято использовать для обозначения некоторых составляющих текста, чаще всего цитат. Рекомендуется по возможности отмечать выделяемый кавычками текст с помощью тегов, а не просто вводить кавычки в тексте. Для этого используются следующие элементы:

`<q>` содержит цитату или кажущуюся цитату — т.е. речь или мысль, отмеченные как принадлежащие другому автору (независимо от того, действительно ли данный отрывок является цитатой); в сюжетно-повествовательных текстах таким образом отмечается прямая речь персонажа или оратора, а в словарях элемент `<q>` можно использовать для обозначения реальных или искусственных примеров использования. Включает следующие атрибуты:

**type** можно использовать для обозначения того, является ли цитируемый отрывок репликой в разговоре или мыслью, а также чтобы более точно охарактеризовать цитату. Предлагаются следующие значения: **spoken** (для обозначения прямой речи, обычно заключаемой в кавычки) и **thought** (для обозначения мыслей, например, внутренних монологов).

**who** идентифицирует произносящего данный текст.

`<mentioned>` отмечает слова или фразы упомянутые, но не использованные.

`<soCalled>` содержит слово или фразу, за которые автор или рассказчик снимает с себя ответственность, что обозначается кавычками или курсивом.

`<gloss>` отмечает слово или фразу, которые представляют собой толкование (глоссу) или определение какого-либо другого слова или фразы. Включает следующие атрибуты:

**target** идентифицирует связанное слово или фразу.

Ниже приведен простой пример цитаты:

```
Few dictionary makers are likely to forget
Dr. Johnson's description of the
lexicographer as <q>a harmless drudge.</q>
```

Чтобы зафиксировать способ, которым напечатана цитата (например, **in-line** (строковая), либо **display** или **block quotation** — блочные цитаты), следует воспользоваться атрибутом **rend**. Кроме того, таким путем можно указать и тип кавычек, в которые заключена цитата.

Чтобы представить прямую речь, прерываемую словами рассказчика, можно просто закончить одну цитату и, после слов автора, начать следующую, как в приведенном ниже примере:

```
<p><q>Who-e debel you?</q> &mdash; he at last said &mdash; <q>you  
no speak-e, damme, I kill-e.</q> And so saying, the lighted  
tomahawk began flourishing about me in the dark.
```

Если важно подчеркнуть то, что два элемента **<q>** вместе воспроизводят одну речь, можно воспользоваться атрибутами связывания **next** и **prev**, которые описаны в разделе 8.3.

Цитаты могут сопровождаться ссылкой на источник или на автора высказывания, для этого используется атрибут **who**, независимо от того, упоминается ли этот источник в тексте, как это имеет место в следующем примере:

```
<q who=Wilson>Spaulding, he came down into the office just this  
day eight weeks with this very paper in his hand, and he  
says: &mdash; <q who=Spaulding>I wish to the Lord, Mr. Wilson, that  
I was a red-headed man.</q></q>
```

Из этого примера также видно, как можно поместить внутри цитаты другую цитату: один персонаж (Wilson) в своей речи цитирует другого (Spaulding).

При создании электронного текста необходимо решить, заменить ли кавычки тегами или, добавив теги, все-таки сохранить кавычки. Если кавычек в тексте не будет, можно использовать атрибут **rend**, чтобы зафиксировать, где они находились в исходном тексте.

Как и при выделении текста, не всегда возможно и не всегда желательно интерпретировать функции кавычек в тексте. В таких случаях можно воспользоваться тегом **<hi rend=quoted>**, чтобы обозначить цитируемый текст, не указывая статус этого текста.

### 6.3 Слова и выражения на иностранных языках

Для обозначения слов и фраз, приведенные на других языках, чем основной текст, можно воспользоваться тегами, причем это можно сделать двумя способами. Если иностранное слово или фраза по какой-либо причине уже отмечены тегами, этот элемент должен включать в себя значение глобального атрибута **lang**, которое и указывает, на каком языке написано данное выражение. Если же иностранные выражения не отмечены тегами, можно воспользоваться элементом **<foreign>**, в котором также используется атрибут **lang**. Например:

```
John has real <foreign lang=fra>savoir-faire</foreign>.  
Have you read <title lang=deu>Die Dreigroschenoper</title>?  
<mentioned lang=fra>Savoir-faire</mentioned> is French for know-how.  
The court issued a writ of <term lang=lat>mandamus</term>.
```

Как ясно из приведенных выше примеров, элемент **<foreign>** следует использовать для обозначения иностранных слов, если не применяются более специальные элементы, например, **<title>**, **<mentioned>** или **<term>**. Глобальный атрибут **lang** можно ввести в любой элемент, чтобы показать, что в нем использован не тот язык, что в остальном тексте.

## 7 Примечания

Любые примечания, которые печатаются в виде сносок, примечаний в конце текста, маргиналий (примечаний на полях) или как-либо по-иному, следует обозначать с помощью одного и того же элемента:

**<note>** содержит примечание или пояснение (аннотацию). Включает следующие атрибуты:

**type** описывает тип примечания.

**resp** указывает, кому принадлежит данное примечание: автору, редактору, переводчику и т.д. Допустимыми являются значения **author**, **editor** и т.п., можно также указать инициалы лица, сделавшего данное примечание.

**place** указывает, в каком месте исходного текста находилось данное примечание. Предусмотрены значения **inline** (строковое), **interlinear** (междустрочное), **left**



(слева) **right** (справа), **foot** (сноска), **end** (в конце текста) для примечаний, которые расположены в основном тексте, между строками, на левом или правом поле, внизу страницы (в нижнем колонтитуле) или в конце главы (тома) соответственно.

**target** означает место в тексте (точку или начало отрывка), к которому относится примечание.

**targetEnd** означает конец отрывка, к которому относится примечание, если оно не вставлено в текст в этой точке.

**anchored** указывает, имеется ли в тексте точное место ссылки на примечание.

Если возможно, текст примечания следует вставлять в основной текст там, где в первый раз появляется идентификатор или метка этого примечания. Это может оказаться невозможным, например, для маргиналий, которые нельзя жестко связать с каким-либо конкретным местом текста. Для простоты можно размещать примечания-маргиналии перед соответствующим абзацем или другим элементом. Примечания можно также поместить в отдельном разделе текста (как в книгах печатают примечания после основного текста) и связать их с соответствующим отрывком с помощью атрибута **target**.

Если необходимо снабдить примечание номером или идентификатором, можно воспользоваться атрибутом **n**. Атрибут **resp** необходимо использовать постоянно, чтобы различать авторские и редакторские примечания, если в произведении имеются примечания обоих типов; в противном случае в заголовке ТЕИ должно быть указано, к какому типу относятся примечания в данном произведении.

Примеры:

```
Collections are ensembles of distinct
entities or objects of any sort.
<note place=foot n=1>
We explain below why we use the uncommon term
<mentioned>collection</mentioned>
instead of the expected <mentioned>set</mentioned>.
Our usage corresponds to the <mentioned>aggregate</mentioned>
of many mathematical writings and to the sense of
<mentioned>class</mentioned> found
in older logical writings. </note> The elements . . .
<lg id=RAM609>
<note place=margin>The curse is finally expiated</note>
<l>And now this spell was snapt: once more</l>
<l>I viewed the ocean green,</l>
<l>And looked far forth, yet little saw</l>
<l>Of what had else been seen &dash;</l>
```

## 8 Перекрестные ссылки и связи

Явные перекрестные ссылки или связи между двумя различными точками текста в одном и том же документе SGML можно закодировать с помощью элементов, описанных в разделе 8.1. Ссылки или связи с элементами некоторых других документов SGML или с частями документов не-SGML можно закодировать с помощью **расширенных указателей ТЕИ**, которые описаны в разделе 8.2. Неявные связи (например, связь между двумя параллельными текстами или между текстом и его интерпретацией) можно закодировать, воспользовавшись атрибутами связывания, которые обсуждаются в разделе 8.3.

### 8.1 Простые перекрестные ссылки

Перекрестную ссылку из одной точки отдельного документа на другую можно закодировать с помощью любого из следующих элементов:

**<ref>** ссылка на другое место текущего документа, выраженная одним или несколькими идентифицируемыми элементами, причем в эти элементы можно ввести дополнительный текст или комментарий.

**<ptr>** ссылка на другое место текущего документа, выраженная одним или несколькими идентифицируемыми элементами.

Эти элементы могут содержать следующие атрибуты:

**target** указывает “пункт назначения” (или “цель”) данного указателя в виде одного или нескольких идентификаторов SGML.

**type** относит данный указатель к определенной категории согласно принятой классификации.

**targType** указывает тип (или типы) элемента, на который “направлен” данный указатель.

**crDate** указывает, когда был создан данный указатель.

**resp** указывает создателя указателя.

Разница между этими двумя элементами заключается в том, что элемент **<ptr>** — это пустой элемент, просто отмечающий точку, связанную с другой точкой, тогда как элемент **<ref>** может содержать также некоторый текст, — обычно текст собственно перекрестной ссылки. Элементом **<ptr>** можно воспользоваться для создания перекрестной ссылки, которая должна быть обозначена какими-либо невербальными способами (например, символом, значком или, в электронных текстах, кнопкой). Этот элемент полезен также в системах создания документов, где верстальщик может корректно создать перекрестную ссылку и в вербальной форме.

Следующие две формы, например, логически эквивалентны (при условии, что точная вербальная форма перекрестной ссылки, представленная в элементе **<ptr>**, задокументирована где-либо):

```
See especially <ref target=SEC12>section 12 on page 34</ref>.
```

```
See especially <ptr target=SEC12>.
```

Значение атрибута **target** должно являться идентификатором SGML в текущем документе SGML. Это означает, что отрывок или фраза, на которую ссылаются, должна иметь идентификатор, а следовательно, должна быть отмечена тегом как элемент некоторого типа. В приведенном ниже примере имеется перекрестная ссылка на элемент **<div1>**:

```
...
see especially <ptr target=SEC12>.
...
<div1 id=SEC12><head>Concerning Identifiers...
...
```

Поскольку атрибут **id** является глобальным, его можно использовать, чтобы сослаться на любой элемент в документе. В следующем примере идентификатором снабжен абзац, на который поэтому можно сослаться таким образом:

```
...
this is discussed in <ref target=pspec>the paragraph on links</ref>
...
<p id=pspec>Links may be made to any kind of element
...
```

Атрибут **targType** можно использовать для указания того, что ссылаются на элемент определенного типа, как это сделано в приведенном ниже примере:

```
...
this is discussed in <ref target=dspec targType='div1 div2'>
the section on links</ref>
```

Эта ссылка окажется ошибочной, если элемент с идентификатором **dspec** не является элементом **<div1>** или **<div2>**. Однако, следует отметить, что нельзя проверить это просто с помощью синтаксического анализатора SGML, поскольку программа синтаксического анализа в SGML указывает только, что элемент **dspec** существует.

Атрибут **type** используется для того, чтобы определить тип связи, представляемой указателем, по любой принятой классификации. Также можно воспользоваться атрибутами **resp** и **crDate**, чтобы указать лицо или организацию, которые создали данную связь, и дату ее создания, как это сделано в следующем примере:

```
...
this is discussed in
<ref type=xref resp=auto crdate=950521 target=dspec targtype='div1 div2'>
the section on links</ref>
```

Эти атрибуты наиболее часто используются в гипертекстовых системах, которые содержат очень много различных указателей, предназначенных для самых разных целей и созданных разнообразными способами.

Иногда “цель” перекрестной ссылки невозможно определить как некоторую конкретную особенность текста и, следовательно, нельзя обозначить как элемент определенного типа. Если целью является просто некоторая точка текущего документа, проще всего обозначить такое назначение с помощью элемента `<anchor>`, который располагается в соответствующем месте. Если цель представляет собой набор слов, не отмеченный никакими другими тегами, можно обозначить этот отрывок с помощью элемента `<seg>`. Свойства этих двух элементов таковы:

- `<anchor>` определяет место или точку в документе, чтобы на них можно было сослаться.
- `<seg>` идентифицирует отрывок или сегмент текста в документе, чтобы на них можно было сослаться. Включает следующие атрибуты:
  - type** определяет категорию сегмента.

В следующем (искусственном) примере элементы `<ref>` использованы для обозначения точек в приведенном тексте, которые необходимо связать каким-либо способом с другими его частями; в первом случае — с точкой, а во втором — с отрывком (набором слов):

```
Returning to <ref target=ABCD>the point where I dozed  
off</ref>, I noticed that <ref target=EFGH>three  
words</ref> had been circled in red by a previous reader
```

При этом способе кодирования необходимо, чтобы элементы с указанными идентификаторами (**ABCD** и **EFGH** в примере выше) встречались еще и в другом месте этого документа. В случае, если предполагается, что в тексте больше не окажется элементов с такими идентификаторами, можно воспользоваться элементами `<anchor>` и `<seg>`:

```
.... <anchor type=bookmark id='ABCD'> ....  
....<seg type=target id='EFGH'> ... </seg> ...
```

Атрибутом **type** следует воспользоваться (как это сделано выше), чтобы указать, с какой именно целью эти элементы общего назначения (а они могут применяться в различных случаях) используются в данном месте данного текста. Другие случаи использования этих элементов обсуждаются в разделе 8.3.

## 8.2 Расширенные указатели

Элементы `<ptr>` и `<ref>` можно использовать только для перекрестных ссылок или связей, “источник” и “цель” которых находятся в пределах одного и того же документа SGML. Кроме того, с помощью этих элементов можно сослаться только на элементы SGML. В этом разделе описаны элементы, не ограниченные такими признаками.

- `<xptr>` определяет указатель (ссылку) на другое место текущего документа или другого (внешнего) документа.
- `<xref>` определяет указатель (ссылку) на другое место текущего документа или другого (внешнего) документа; возможно, с дополнительным текстом или комментарием.

Кроме атрибутов указателей, уже описанных выше в разделе 8.1, эти элементы могут иметь дополнительные атрибуты, которые используются для определения “цели” перекрестной ссылки или связи вместо атрибута **target**:

- doc** указывает документ, в котором следует искать требуемое место; значение по умолчанию — текущий документ.
- from** отмечает начало пункта назначения (“цели”) указателя, представленного в виде выражения, соответствующего синтаксису расширенных указателей ТЕИ; значение по умолчанию — весь документ, указанный в атрибуте **doc**.
- to** отмечает конец пункта назначения (“цели”) указателя, представленного в виде выражения, соответствующего синтаксису расширенных указателей ТЕИ; используется только в том случае, если определен атрибут **from**.

Полное описание языка, который используется для обозначения целей расширенных указателей ТЕИ, не входит в задачу данного документа; здесь перечислены только некоторые из наиболее часто используемых функций. Подробности можно найти в полном варианте руководстве.

Элемент `<xptr>` (или `<xref>`) может указывать на какой-либо другой документ в целом, для этого нужно просто указать название нужного объекта в качестве значения атрибута **doc**, как это сделано в следующем примере:

```
see <xref doc=P3>The TEI Guidelines, passim</xref>
```

В этом примере подразумевается, что ранее описана некоторая система или общедоступный объект с именем **P3**. Это описание может находиться в файле расширения **litemods.ent** или может быть сделано другим способом, присущим тому конкретному авторскому программному обеспечению SGML, которое используется в данном случае (эти вопросы рассмотрены в разделе 15).

Атрибут **from** используется для того, чтобы определить конкретное место в документе, который указан в атрибуте **doc**. В таких спецификациях используется специальный язык, который называется **Синтаксис расширенных указателей TEI** (TEI extended pointer syntax); здесь приводятся только некоторые сведения о нем. С помощью этого языка местонахождение в документе определяется как ряд “шагов” (steps), каждый из которых идентифицирует определенную часть документа, часто в терминах местонахождений, определенных предыдущим шагом. Например, чтобы указать на третье предложение во втором абзаце главы 2, следует выбрать вторую главу в первом шаге, второй абзац во втором, и третье предложение в последнем шаге. Шаг можно определить в терминах понятий SGML (таких как **parent**, **descendent**, **preceding** и т.д.) или, более свободно, в терминах структуры текста, позиций слов или символов. Можно также использовать иную (не-SGML) систему обозначений или указать место в графическом объекте с помощью связанной с ним системы координат.

Атрибуты **from** и **to** используют одну и ту же систему обозначений. Каждый указывает на некоторый отрывок в документе-цели; расширенный указатель как целое указывает на раздел, который начинается с отрывка, указанного в атрибуте **from** и заканчивается текстом, указанным в атрибуте **to**.

Первый шаг в пути к месту назначения часто будет состоять в указании идентификатора некоторого элемента в целевом документе, как показано в следующем примере:

```
<xptr doc=P3 from='id (SA) '>
```

Таким образом, в объекте **P3** выбирается любой элемент с идентификатором **SA**. Если необходимо более подробно определить цель, можно ввести следующие шаги. Можно воспользоваться следующими ключевыми словами для выбора других элементов, причем эти элементы определяются по их соотношению с упомянутым:

**child** элементы, содержащиеся в данном элементе.

**ancestor** элементы, содержащие в себе (прямо или косвенно) данный элемент.

**previous** элементы того же уровня в иерархии, что и данный, но предшествующие ему в документе.

**next** элементы того же уровня в иерархии, что и данный, но следующие за ним в документе

**preceding** элементы в документе, начинающиеся перед данным элементом, без учета их уровня в иерархии.

**following** элементы в документе, начинающиеся после данного элемента, без учета их уровня в иерархии.

Каждое из этих ключевых слов обозначает определенный набор элементов (набор порожденных элементов, набор элементов-предков, набор предшествующих элементов одного уровня и т.д.); чтобы определить указание на какой-либо конкретный элемент подобного множества, можно дополнить ключевое слово заключенным в скобки списком, содержащим:

- положительное или отрицательное число, означающее, на какой из обнаруженных нескольких элементов направлен указатель (+1 обозначает первый обнаруженный элемент, начиная с текущего места, а -1 обозначает последний элемент), либо ключевое слово **all**, которое означает, что указатель направлен на все элементы из обнаруженного множества;
- идентификатор, указывающий тип искомого элемента, или звездочку, которая означает, что элемент может быть любого типа;
- набор имен и значений атрибутов, означающий, что выбранный элемент должен иметь атрибуты с указанными именами и значениями, если они приведены.

Возвращаясь к приведенному выше примеру, видим, что в следующей ссылке будет выбран третий элемент **<p>**, непосредственно содержащийся в любом элементе с идентификатором **SA**:

```
<xptr doc=P3 from='id (SA) child (3 p) '>
```

Аналогично, если предположить, что объект **P3** в действительности является SGML-формой “Принципов TEI”, то в приведенной ниже ссылке выбран раздел 14.2.2 этой публикации, в котором (как это и имеет место) дано формальное определение синтаксиса расширенных указателей:

For full details, see

```
<ref doc=P3 from='id (SA) child (2 div2) child (2 div3) '>
```

TEI Extended pointer syntax definition

```
</ref>
```

Обычно позиция перекрестной ссылки точно определяется атрибутом **from**. Для некоторых документов, однако, может оказаться более удобным определить и начальную, и конечную позицию. Как отмечалось выше, для этой цели служит атрибут **to**. Например, выражение

```
<xptr doc=P1 from='id (xyz)' to='id (abc) '>
```

представляет собой расширенный указатель, “цель” которого — это последовательность, начало которой совпадает с началом в документе **P1** элемента любого типа с идентификатором **XYZ**, а конец — с окончанием (в том же документе) любого элемента с идентификатором **ABC**. Все элементы между этими двумя также включаются, независимо от структуры; указатель окажется ошибочным, если конец элемента **ABC** предшествует началу элемента **XYZ**.

Используя этот синтаксис, легко построить очень сложные спецификации. Например, по следующей ссылке будет выбран самый последний элемент **<head>**, у которого атрибут **lang** имеет значение **LAT**, и который находится перед началом элемента с идентификатором **SA**:

```
<xptr doc=P3 from='id (SA) preceding (1 head lang lat) '>
```

Если для атрибута **doc** не определено значение, это означает текущий документ. Таким образом, следующие ссылки семантически одинаковы. Они обе указывают на элемент с идентификатором **X1**, находящийся в текущем документе:

```
<ptr target=X1>  
<xptr from='id (X1) '>
```

### 8.3 Атрибуты связывания

Следующие атрибуты **связывания** определены для каждого элемента в TEI Lite DTD:

**ana** связывает элемент и его интерпретацию.

**corresp** связывает элемент с одним или несколькими соответствующими элементами.

**next** связывает элемент со следующим элементом в совокупности (элементов).

**prev** связывает элемент с предыдущим элементом в совокупности (элементов).

Атрибут **ana** (анализ) следует использовать в тех случаях, когда где-либо в пределах одного документа определено множество абстрактных анализов или интерпретаций, эти вопросы обсуждаются далее в разделе 16. Например, лингвистический анализ предложения “John loves Nancy” (Джон любит Нэнси) можно закодировать следующим образом:

```
<seg type=sentence ana=SVO>  
  <seg type=lex ana=NP1>John</seg>  
  <seg type=lex ana=VVI>loves</seg>  
  <seg type=lex ana=NP1>Nancy</seg>  
</seg>
```

Такое кодирование подразумевает, что где-либо в документе существуют элементы с идентификаторами **SVO**, **NP1** и **VVI**, в которых объясняется значение этих конкретных кодов. Следует обратить внимание на использование элемента **<seg>** для обозначения отдельных компонентов анализа, различаемых с помощью атрибута **type**.

Атрибут **corresp** (соответствие) обеспечивает простой способ представления некоторых форм соответствия между двумя элементами в тексте. Например, в “многоязычном” тексте (т.е. в тексте, приведенном на нескольких языках) можно воспользоваться этим атрибутом, чтобы связать эквиваленты перевода, как это сделано в следующем примере:

```
<seg lang=FRA id=FR1 corresp=EN1>Jean aime Nancy</seg>  
<seg lang=ENG id=EN1 corresp=FR1>John loves Nancy</seg>
```

Этот способ можно использовать в различных случаях. Так, в следующем примере он использован для представления анафорных соответствий между “the show” и “Shirley” и “NBC” и “the network”:

```
<p><title id=shirley>Shirley</title>, which made  
its Friday night debut only a month ago, was  
not listed on <name id=nbc>NBC</name>'s new schedule,  
although <seg id=network corresp=nbc>the network</seg>  
says <seg id=show corresp=shirley>the show</seg>  
still is being considered.
```

Использование атрибутов **next** и **prev** — это простой способ связать друг с другом компоненты прерывающегося элемента; ниже приведен пример такого применения этих атрибутов:

```
<q id=Q1a next=Q11b>Who-e debel you?</q>
&mdash he at last said &mdash
<q id=Q1b prev=Q1a>you no speak-e,
damme, I kill-e.</q> And so saying,
the lighted tomahawk began flourishing
about me in the dark.
```

## 9 Редакторская правка

Процесс кодирования электронного текста имеет много общего с процессом редактирования рукописи или другого текста для печатной публикации. В обоих случаях добросовестный редактор желает зафиксировать и исходное состояние источника, и всю редакторскую правку или другие изменения, внесенные в текст. В этом и в следующем разделах описаны элементы, которые обеспечивают некоторые средства для решения таких задач.

Следующую пару элементов можно использовать для обозначения **исправлений**, т.е. изменений, внесенных редактором в тех местах, где по его мнению, в оригинале имеется ошибка:

- <corr>** содержит правильную форму отрывка, который в тексте приведен с явными ошибками. Предусмотрены следующие атрибуты:
- sic** дает исходную форму явной ошибки в тексте.
  - resp** означает редактора или публикатора, предложившего исправление, которое зафиксировано как значение элемента **<corr>**.
  - cert** обозначает степень уверенности в необходимости исправления, которое зафиксировано как значение элемента **<corr>**.
- <sic>** содержит текст, который следует воспроизвести неизменным, несмотря на его явную некорректность или неточность. Предусмотрены следующие атрибуты:
- corr** дает исправление явной ошибки в тексте.
  - resp** означает редактора или корректора, предложившего данное исправление.
  - cert** обозначает степень уверенности в необходимости данного исправления.

Описанной ниже парой элементов можно воспользоваться, чтобы отметить **нормализацию**, т.е. редакторскую правку, которая внесена с целью сделать текст более логичным или современным:

- <orig>** содержит исходное прочтение, а исправленный вариант приведен в виде значения атрибута. Предусмотрены следующие атрибуты:
- reg** дает исправленный (нормализованный) вариант текста.
  - resp** определяет лицо, ответственное за исправление слова или фразы.
- <reg>** содержит прочтение, исправленное или нормализованное по определенным признакам. Предусмотрены следующие атрибуты:
- orig** дает неисправленный вариант текста, т.е. текст, приведенный в источнике.
  - resp** определяет лицо, ответственное за исправление слова или фразы.

Например, прочтение

```
... for his nose was as sharp as a pen and a' table of green feelds
рассматривается редактором Гиффордом (Gifford) как требующее следующих примечаний: (1) ошибочное
использование слова table вместо babbled и (2) нестандартное написание a' и feelds вместо he и fields.
Замечания Гиффорда об исправлении текста можно закодировать следующим образом:
```

```
... for his nose was as sharp as a pen and <reg sic="a'">he</reg>
<corr sic='table' ed=Gifford>babbl'd</corr> of green
<reg sic='feelds'>fields</reg>
```

## 10 Пропуски, удаления и дополнения

Кроме исправления или нормализации слов и фраз, редакторы и публикаторы могут предложить вставить пропущенные материалы, не публиковать некоторые материалы или ввести материалы,

вычеркнутые из оригинала. Кроме того, некоторые материалы могут оказаться исключительно сложными для транскрипции, поскольку их непросто разместить на странице. Справиться со всеми перечисленными задачами позволяют следующие элементы:

- <add>** содержит буквы, слова или фразы, вставленные в текст автором, переписчиком, составителем аннотации или корректором. Предусмотрены следующие атрибуты:
- place** если дополнение вписано в рукописный текст, указывает, где размещен этот дополнительный текст. Допустимы значения **inline** (в строке), **supralinear** (над строкой), **infralinear** (под строкой), **left** (на левом поле), **right** (на правом поле), **top** (вверху страницы), **bottom** (внизу страницы) и т.д.
- <gap>** указывает точку, где при переписке были пропущены материалы, либо по редакторским причинам, описанным в заголовке ТЕИ, как часть принятой практики выбора, либо из-за того, что эти материалы неудобочитаемы или невняты. Предусмотрены следующие атрибуты:
- desc** дает описание опущенного текста.
  - resp** указывает редактора, публикатора или кодировщика, принявшего решение пропустить этот текст, и, следовательно, применить тег **<gap>**.
- <del>** содержит букву, слово или отрывок, удаленные, отмеченные как удаленные, либо другим способом обозначенные как излишние или неверные в рукописном тексте автором, переписчиком, составителем аннотации или корректором. Предусмотрены следующие атрибуты:
- type** указывает тип удаленного материала в соответствии с принятой классификацией.
  - status** может использоваться для обозначения ошибочных удалений, например, если вычеркнут излишний текст или наоборот, подлежащее удалению вычеркнуто не до конца.
  - hand** определяет исполнителя, который удалил материал.
- <unclear>** содержит слово, фразу или отрывок, которые нельзя расшифровать с уверенностью из-за неудобочитаемости или невнятности оригинала. Предусмотрены следующие атрибуты:
- reason** указывает причину, по которой данный материал сложен для расшифровки.
  - resp** указывает лицо, выполнившее расшифровку (и ответственное за нее) буквы, слова или отрывка, содержащегося в элементе **<unclear>**.

Эти элементы могут использоваться для регистрации изменений, внесенных редактором, публикатором или (для рукописных материалов) автором или переписчиком. Например, если оригинал электронного текста читается как

```
The following elements are provided for  
for simple editorial interventions.
```

может оказаться необходимым не только исправить очевидную ошибку, но в то же время зарегистрировать удаление лишнего второго **for**, т.е. закодировать текст следующим образом:

```
The following elements are provided for  
<del hand=LB>for</del> simple editorial interventions.
```

Значение **LB** атрибута **hand** указывает, что исполнитель “LB” удалил дублирующийся предлог **for**. Если оригинал читался как

```
The following elements provided for  
for simple editorial interventions.
```

(т.е. если случайно оказался также пропущен глагол), исправленный текст может выглядеть следующим образом:

```
The following elements <add hand=LB>are</add> provided for  
<del hand=LB>for</del> simple editorial interventions.
```

Значение **LB** атрибута **hand** указывает, что исполнитель “LB” добавил пропущенный глагол и удалил дублирующийся предлог **for**.

Использование этих элементов не ограничено только редакторской правкой; с их помощью можно также зафиксировать авторскую правку в рукописях. Если автор сначала написал в рукописи "How it galls

me, what a galling shadow" (Как это раздражает меня, что за раздражающий полумрак), а затем вычеркнул слово **galls** (раздражает) и вместо него вставил **dogs** (преследует), это можно закодировать следующим образом:

```
How it <del hand=DHL type=overstrike>galls</del>
<add hand=DHL place=supralinear>dogs</add> me,
what a galling shadow
```

Аналогично, элементы **<unclear>** и **<gap>** можно использовать в сочетании, чтобы указать пропуск невнятных материалов; в приведенном примере также показано, как использовать элемент **<add>** для предположительного исправления:

```
One hundred & twenty good regulars joined to me
<unclear><gap reason='indecipherable'></unclear>
& instantly, would aid me signally <add hand=ed>in?</add>
an enterprise against Wilmington.
```

Элементом **<del>** отмечаются материалы, которые включены в электронный текст, несмотря на то, что они отмечены как удаленные, в то время как элемент **<gap>** отмечает местонахождение материалов, которые опускаются в электронном тексте, независимо от того, разборчивы они или нет. Например, в совокупности текстов на каком-либо одном языке можно опустить длинные цитаты на иностранных языках:

```
<p> ... An example of a list appearing in a fief ledger of
<name type=place>Koldinghus</name> <date>1611/12</date>
is given below. It shows cash income from a sale of
honey.</p>
<q><gap desc='quotation from ledger'
reason='in Danish'></q>
<p>A description of the overall structure of the account is
once again ... </p>
```

В других совокупностях (особенно в тех, что были созданы до начала широкого применения сканеров) систематически опускаются рисунки и формулы:

```
<p>At the bottom of your screen below the mode line is the
<term>minibuffer</term>. This is the area where Emacs
echoes the commands you enter and where you specify filenames
for Emacs to find, values for search and replace,
and so on.
<gap desc='diagram of Emacs screen' reason='graphic'>
</p>
```

## 11 Имена, даты, числа и сокращения

В системе TEI определены элементы для большого числа “подобных данным” характеристик, которые могут появляться почти в любом месте текстов практически всех типов. Эти характеристики могут представлять особый интерес для целого ряда дисциплин; они все связаны с объектами, внешними по отношению к собственно тексту. Такими объектами могут быть имена людей, названия местностей, числа, даты. Они также создают определенные проблемы для многих приложений естественно-языковой обработки (NLP-приложений), поскольку существует множество способов представления подобных объектов в тексте. Описанные здесь элементы делают эти объекты явными и, таким образом, уменьшают сложность обработки текстов, содержащих их.

### 11.1 Имена и называющие строки

**Называющая строка** — это фраза, представляющая собой ссылку на определенное лицо, место, объект и т.п. Для обозначения таких строк предусмотрено два элемента:

**<rs>** содержит имя общего назначения или называющую строку. Предусмотрены следующие атрибуты:

**type** более конкретно определяет объект, упомянутый в называющей строке. Может принимать значения **person** (лицо), **place** (место), **ship** (корабль), **element** (элемент) и т.д.

**<name>** содержит имя собственное, которое может представлять собой фразу. Предусмотрены следующие атрибуты:



**type** указывает тип объекта, носящего данное имя.

Атрибут **type** используется для разделения (например), имен людей, названий местностей и названий организаций, где это возможно:

```
<q>My dear <rs type=person>Mr. Bennet</rs>, </q>
said his lady to him one day, <q>have you heard that
<rs type=place>Netherfield Park</rs> is let
at last?</q>
```

```
It being one of the principles of the
<rs type=organization>Circumlocution Office</rs> never,
on any account whatsoever, to give a straightforward answer,
<rs type=person>Mr Barnacle</rs> said, <q>Possibly.</q>
```

Как показано в следующем примере, элемент **<rs>** можно использовать для любой ссылки на лицо, место и т.д., причем не обязательно, чтобы в этой ссылке было имя собственное.

```
<q>My dear <rs type=person>Mr. Bennet</rs>,</q>
said <rs type=person>his lady</rs> to him
one day ...
```

Элемент **<name>**, напротив, предназначен только для определенного типа называющих строк, которые содержат только имена собственные; этот элемент можно использовать синонимично с элементом **<rs>**, или вставить его в элемент **<rs>**, если называющая строка содержит и имена собственные, и другие существительные.

Просто обозначить тегом что-либо как имя обычно недостаточно для автоматического перевода личных имен в канонические формы, что обычно требуется для справочных целей. Имя, как оно появляется в тексте, может быть неверно написанным, неполным или неправильным. Более того, префиксы перед именами, такие как **van** или **de la** могут входить или не входить в принятую для ссылок форму имен, что зависит от языка и страны происхождения носителя данного имени.

Для преодоления таких трудностей в этих и подобных им элементах можно воспользоваться следующими атрибутами:

**key** обеспечивает альтернативный идентификатор для объекта, носящего данное имя, подобный ключу записи в базе данных.

**reg** дает нормализованную или исправленную форму использованного имени.

Атрибут **key** полезен в качестве средства, позволяющего собрать вместе все ссылки на данное лицо или место, разбросанные по всему документу:

```
<q>My dear <rs type=person key=BENM1>Mr. Bennet</rs>,
</q> said <rs type=person key=BENM2>his lady</rs>
to him one day, <q>have you heard that
<rs type=place key=NETP1>Netherfield Park</rs>
is let at last?</q>
```

Этот способ применения следует отличать от использования атрибута **reg** (нормализация), который позволяет обозначить стандартную форму называющей строки, как показано в приведенном ниже примере:

```
<name type=person key=WADLM1 reg='de la Mare, Walter'>
  Walter de la Mare
</name>
was born at
<name key=Ch1 type=place>Charlton</name>,
in <name key=KT1 type=county>Kent</name>, in 1873.
```

Более детальная разметка имен, т.е. присвоение тегов компонентам имен собственных, также возможна, для этого используется дополнительный набор тегов для имен и дат.

## 11.2 Даты и время

Теги для более детального кодирования времени и даты включают в себя следующее:

**<date>** содержит дату в любом формате. Допускаются следующие атрибуты:

**calendar** означает систему счисления или календарь, которому соответствует дата.

**value** дает значение даты в какой-либо стандартной форме, обычно в формате

ГГГГ-ММ-ДД.

**<time>** содержит фразу, определяющую время суток в любом формате.

**value** дает значение времени в стандартной форме.

Атрибут **value** определяет нормализованную форму для даты или времени, используя какой либо признанный формат, например, ISO 8601. Неполные даты или время (например, “1990”, “сентябрь 1990”, “около двенадцати”) обычно можно выразить, просто опустив часть полного значения; с другой стороны, неточные даты или время (например, “в начале августа”, “где-то между десятью и двенадцатью”) можно выразить через диапазон дат или интервал времени. Если одна из границ такого диапазона известна точно (например, “еще до 1230 года”, “спустя несколько дней после Хэллоуина”), можно воспользоваться атрибутом **exact**, чтобы отметить этот факт.

Примеры:

```
<date value='1980-02-21'>21 Feb 1980</date>
```

```
<date value='1990'>1990</date>
```

```
<date value='1990-09'>September 1990</date>
```

```
Given on the <date value='1977-06-12'>Twelfth Day of June  
in the Year of Our Lord One Thousand Nine Hundred and  
Seventy-seven of the Republic the Two Hundredth and first  
and of the University the Eighty-Sixth. </date>
```

```
<l>pecially when it's nine below zero
```

```
<l>and <time value='15:00'>three o'clock in the afternoon</time>
```

### 11.3 Числа

Числа могут быть записаны как буквами, так и цифрами (двадцать один, xxi, 21), и их представление зависит от языка (так, по-английски следует писать **5th**, тогда как по-гречески — **5**; число, в английском тексте представленное как **123,456.78**, во французском будет выглядеть как **123.456,78**). При естественно-языковой обработке и в приложениях машинного перевода часто бывает полезным установить различие между числами и более “лексическими” частями текста. В других приложениях важно записать значение числа в стандартной системе. Элемент **<num>** обеспечивает эту возможность:

**<num>** содержит число, записанное в любой форме. Предусмотрены следующие атрибуты:

**type** указывает тип числового значения. Предусмотрены следующие значения:

**fraction** (дробь), **ordinal** (для порядковых номеров, например, “21st”), **percentage** (процент) и **cardinal** (абсолютное число, например, “21”, “21.5” и т.п.).

**value** содержит значение числа в форме, стандартной для приложения.

Например:

```
<num value='33'>xxxiii</num>
```

```
<num type=cardinal value='21'>twenty-one</num>
```

```
<num type=percentage value='10'>ten percent</num>
```

```
<num type=percentage value='10'>10%</num>
```

```
<num type=ordinal value='5'>5th</num>
```

### 11.4 Сокращения и их полные формы

Подобно именам, датам и числам, сокращения можно записывать как в собственно сокращенной, так и в полной форме; их можно оставить необозначенными или закодировать с помощью следующего элемента:

**<abbr>** содержит сокращение любого типа. Предусмотрены следующие атрибуты:

**expand** содержит полную форму сокращения.

**type** позволяет указать тип сокращения согласно принятой классификации.

Предусмотрены следующие значения: **contraction**, **suspension**, **brevigraph**, **superscription** и **acronym**. Атрибуту **type** можно также присвоить такие значения, как **title** (для названий в адресах), **geographic**, **organization** и т.д., чтобы описать объект, на который ссылаются.

Элемент **<abbr>** также полезен в качестве средства для обозначения “полу-лексических” единиц, например, акронимов или жаргонизмов:

We can sum up the above discussion as follows: the identity of a **<abbr>CC</abbr>** is defined by that calibration of values which motivates the elements of its **<abbr>GSP</abbr>**;

Every manufacturer of **<abbr>3GL</abbr>** or **<abbr>4GL</abbr>** languages is currently nailing on **<abbr>OOP</abbr>** extensions

Атрибут **type** можно использовать для выделения типов сокращений по их функции, а атрибут **expn** используется для определения полной формы сокращений:

```
<name><abbr type=title expn='Doctor'>Dr.</abbr>
<abbr type=initial expn='Marilyn'>M.</abbr>
Deegan</name>
is the Director of the
<abbr expn='Computers in Teaching Initiative' type=acronym>
CTI</abbr> Centre for Textual Studies.
```

Этот элемент особенно полезен при работе с рукописными материалами, в которых сокращения используются очень часто.

## 11.5 Адреса

Элемент **<address>** используется для обозначения почтовых адресов всех видов. Он содержит один или несколько элементов **<addrLine>**, по одному для каждой строки адреса.

**address** содержит почтовый или другой адрес, например, издателя, организации или частного лица.

**addrLine** содержит одну строку почтового или другого адреса.

Вот простой пример:

```
<address>
<addrLine>Computer Center (M/C 135)</addrLine>
<addrLine>1940 W. Taylor, Room 124</addrLine>
<addrLine>Chicago, IL 60612-7352</addrLine>
<addrLine>U.S.A.</addrLine>
</address>
```

Можно выделить в адресе еще более мелкие части, воспользовавшись элементом **<name>**, который описан выше (раздел 11.1).

```
<address>
<addrLine>Computer Center (M/C 135)</addrLine>
<addrLine>1940 W. Taylor, Room 124</addrLine>
<addrLine><name type=city>Chicago</name>, IL 60612-7352</addrLine>
<addrLine><name type=country>USA</name></addrLine>
</address>
```

## 12 Списки

Элемент **<list>** используется для обозначения **списков** любого типа. Список представляет собой последовательность текстовых единиц, которая может быть упорядоченной, неупорядоченной или глоссарием. Перед каждой единицей может располагаться ее метка (в случае глоссария такой меткой является определяемый термин):

**<list>** содержит любую последовательность единиц, составляющих список. Предусмотрены следующие атрибуты:

**type** описывает форму списка. Предусмотрены следующие значения: **ordered**, **bulleted** (соответственно, для списков, пункты которых пронумерованы или отмечены буквами, и для списков, в которых пункты отмечены значком, например, жирной точкой), **gloss** (для списков, состоящих из набора технических терминов, каждый из которых отмечен элементом **<label>** и сопровождается толкованием или определением, обозначенным с помощью элемента **<item>**) и **simple** (для списков, пункты которых не пронумерованы или не отмечены другим способом).

**<item>** содержит один компонент списка.

**<label>** содержит метку, связанную с компонентом списка; в глоссариях отмечает определяемый термин.

Отдельные пункты списка отмечаются с помощью тега **<item>**. Перед первым из элементов **<item>** можно (но не обязательно) разместить элемент **<head>**, в котором содержится заголовок списка. Нумерация в списке может быть опущена (если предполагается, что список будет перестраиваться), определена с помощью атрибута **n**, который в этом случае указывается в каждом пункте списка, или (редко) отмечена тегом как содержание с помощью элемента **<label>**. Таким образом, следующие три примера эквивалентны:

```
<list>
<head>A short list</head>
<item>First item in list.</item>
<item>Second item in list.</item>
<item>Third item in list.</item>
</list>
```

```
<list>
<head>A short list</head>
<item n=1>First item in list.</item>
<item n=2>Second item in list.</item>
<item n=3>Third item in list.</item>
</list>
```

```
<list>
<head>A short list</head>
<label>1</label><item>First item in list.</item>
<label>2</label><item>Second item in list.</item>
<label>3</label><item>Third item in list.</item>
</list>
```

В одном и том же списке нельзя использовать разные стили одновременно.

Простая таблица из двух столбцов может рассматриваться как **глоссарий**, отмеченный тегом **<list type=gloss>**. Здесь каждый пункт списка состоит из термина и толкования (глосса), которые обозначены, соответственно, элементами **<label>** и **<item>**. Эти элементы соответствуют элементам **<term>** и **<gloss>**, которые могут располагаться в любом месте написанного прозой текста.

```
<list type=gloss>
<head>Vocabulary</head>
<label lang=enm>nu</label>          <item>now</item>
<label lang=enm>lhude</label>      <item>loudly</item>
<label lang=enm>bloweth</label>    <item>blooms</item>
<label lang=enm>med</label>        <item>meadow</item>
<label lang=enm>wude</label>       <item>wood</item>
<label lang=enm>awe</label>        <item>ewe</item>
<label lang=enm>lhouth</label>     <item>lows</item>
<label lang=enm>sterteth</label>   <item>bounds, frisks</item>
<label lang=enm>verteth</label>   <item lang=lat>pedit</item>
<label lang=enm>murie</label>     <item>merrily</item>
<label lang=enm>swik</label>      <item>cease</item>
<label lang=enm>naver</label>     <item>never</item>
</list>
```

В случае, если пункт списка имеет более сложную внутреннюю структуру, предпочтительнее рассматривать список как **таблицу**, поскольку для разметки таблиц определены специальные теги в дополнительном наборе тегов TEI.

Хотя, безусловно, пункты списков любого типа могут иметь как угодно сложную структуру. Так, в приведенном ниже примере глоссарий состоит из двух пунктов, каждый из которых представляет собой простой список:

```
<list type=gloss><label>EVIL</label>
<item><list type=simple>
  <item>I am cast upon a horrible desolate island, void
    of all hope of recovery.</item>
```

```

<item>I am singled out and separated as it were from
    all the world to be miserable.</item>
<item>I am divided from mankind &mdash a solitaire; one
    banished from human society.</item>
</list> <!-- end of first nested list --></item>
<label>GOOD</label>
<item><list type=simple>
    <item>But I am alive; and not drowned, as all my
    ship's company were.</item>
    <item>But I am singled out, too, from all the ship's
        crew, to be spared from death...</item>
    <item>But I am not starved, and perishing on a barren place,
        affording no sustenances....</item>
</list><!-- end of second nested list --></item>
</list><!-- end of glossary list -->

```

Список совершенно необязательно выводить на экран в формате списка. Например:

```

On those remote pages it is written that animals are
divided into <list rend="run-on"><item n='a'>those that belong to the
Emperor,<itemn='b'> embalmed ones, <item n='c'> those
that are trained, <item n='d'> suckling pigs, <item n='e'>
mermaids, <item n='f'> fabulous ones, <item n='g'> stray
dogs, <item n='h'> those that are included in this
classification, <item n='i'> those that tremble as if they
were mad, <item n='j'> innumerable ones, <item n='k'> those
drawn with a very fine camel's-hair brush, <item n='l'>
others, <item n='m'> those that have just broken a flower
vase, <item n='n'> those that resemble flies from a
distance.</list>

```

Списки литературы (библиография) должны иметь тег, использующий элемент **<listBibl>**. Этот элемент описан в следующем разделе.

### 13 Библиографические ссылки

Если в тексте встречаются библиографические ссылки, полезно выделить их в явной форме хотя бы для того, чтобы при печати текста они были сформатированы надлежащим образом, а также для исследовательских целей. Для этого предназначен элемент **<bibl>**:

**<bibl>** содержит библиографическую ссылку произвольной структуры, компоненты которой можно либо явно отметить тегами, либо не выделять.

При выделении компонентов библиографической ссылки следует пользоваться перечисленными ниже элементами. В большинстве случаев полезно отметить по крайней мере те фрагменты (например, названия статей, книг и журналов), для которых необходимо специальное форматирование. Для случаев, когда таким подробностям уделяется особое внимание, предусмотрены дополнительные элементы.

**<author>** в библиографической ссылке содержит имя автора (авторов) произведения, персональное или имя коллективного автора; это первичные сведения **об ответственности** для любой библиографической единицы.

**<biblScope>** определяет степень подробности (“объем”) библиографической ссылки, например, указаны ли в ней только номера страниц или приведены также названия подразделов большого произведения.

**<date>** содержит дату в любом формате.

**<editor>** вторичные сведения **об ответственности** для библиографической единицы, например, имя конкретного лица либо название учреждения или организации (возможно, нескольких), выступающих в роли редактора, составителя, переводчика и т.п. Для этого элемента предусмотрены следующие атрибуты:

**role** определяет тип интеллектуальной ответственности. Предусмотрены значения **translator** (переводчик), **compiler** (составитель), **illustrator** (иллюстратор) и т.д. По умолчанию принимается значение **editor** (редактор).

- <imprint>** содержит информацию, связанную с публикацией или распространением библиографической единицы.
- <publisher>** содержит название организации, ответственной за публикацию или распространение библиографической единицы.
- <pubPlace>** содержит название места, где издана библиографическая единица.
- <series>** содержит информацию о серии, в которой выходит данная книга или другая библиографическая единица.
- <title>** содержит название произведения, будь то статья, книга, журнал или серия, включая все альтернативные названия или подзаголовки. Предусмотрены следующие атрибуты:
- type** указывает тип названия по принятой классификации, например, **main** (основное название), **subordinate** (подзаголовков) и т.д.
- level** указывает **библиографический класс** названия. Допустимые значения этого атрибута описаны в разделе 6.1.

Например, следующее примечание редактора: He was a member of Parliament for Warwickshire in 1445, and died March 14, 1470 (according to Kittredge, *Harvard Studies* 5. 88ff) [Он был членом Парламента от графства Варвикшир в 1445 г. и умер 24 марта 1470 г. (см. Киттредж, Гарвардские Записки, 5. 88ff)] записывается в следующей форме:

```
He was a member of Parliament for Warwickshire in 1445, and died
March 14, 1470 (according to <bibl><author>Kittredge</author>,
<title>Harvard Studies</title> <biblescope>5. 88ff</biblescope></bibl>).
```

Списки библиографических ссылок (библиографии) определяются элементом **<listBibl>**, который может содержать несколько элементов **<bibl>**. Пример такого списка приведен в разделе 22.

## 14 Таблицы

Работа с таблицами является сложной задачей для любой системы обработки текстов, но хотя бы простые таблицы необходимы в таком множестве текстов, что даже в представленном здесь упрощенном наборе тегов TEI имеются средства для разметки таблиц. Для этой цели служат следующие элементы:

- <table>** содержит текст, выводимый в виде таблицы, т.е. состоящий из строк и столбцов. Предусмотрены следующие атрибуты:
- rows** указывает количество строк в таблице.
- cols** указывает количество столбцов в каждой строке таблицы.
- <row>** содержит одну строку таблицы. Предусмотрены следующие атрибуты:
- role** указывает тип информации, хранимой в ячейках этой строки. Допустимы следующие значения: **label** для меток или описательной информации и **data** для реальных значений данных.
- <cell>** содержит одну ячейку таблицы. Предусмотрены следующие атрибуты:
- role** указывает тип информации, хранимой в ячейке. Допустимы следующие значения **label** для меток или описательной информации и **data** для реальных значений данных.
- cols** указывает число столбцов, занимаемых этой ячейкой.
- rows** указывает число строк, занимаемых этой ячейкой.

Например, Дефо (Defoe) следующим образом воспользовался таблицами смертности в *Дневнике года чумы* (Journal of the Plague Year), чтобы показать подъем и спад эпидемии:

```
<p>It was indeed coming on amain, for the burials that
same week were in the next adjoining parishes thus:&mdash;
<table rows=5 cols=4>
<row role='data'>
<cell role='label'>St. Leonard's, Shoreditch</cell>
  <cell>64</cell> <cell>84</cell> <cell>119</cell></row>
<cell role='label'>St. Botolph's, Bishopsgate</row>
  <cell>65</cell> <cell>105</cell> <cell>116</cell></row>
<cell role='label'>St. Giles's, Cripplegate</row>
  <cell>213</cell> <cell>421</cell> <cell>554</cell></row>
```

```
</table>
<p>This shutting up of houses was at first counted a very cruel
and unchristian method, and the poor people so confined made
bitter lamentations. ... </p>
```

## 15 Рисунки и графика

Не все компоненты документа обязательно являются текстовыми. Самые простые тексты часто содержат диаграммы или иллюстрации, не говоря уже о документах, в которых изображение и текст переплетены так, что их невозможно отделить друг от друга, или об электронных ресурсах, в которых эти составляющие взаимно дополняют друг друга.

Кодировщик может легко зафиксировать наличие графики в тексте, возможно, даже с кратким ее описанием, воспользовавшись элементами, описанными в данном разделе. Этими же элементами можно воспользоваться для введения цифровых версий графики в электронный документ.

**<figure>** отмечает место, где в документ будет вставлена графика. Предусмотрены следующие атрибуты:

**entity** название ранее определенного системного объекта, содержащего цифровую версию вводимой графики.

**<figDesc>** содержит текст описания, касающегося появления или содержания графики; это описание используется в тех случаях, если об изображении только сообщается, но оно не выводится.

Любую текстовую информацию, сопровождающую графическое изображение, например, заголовок и/или пояснительные надписи, можно включить в сам элемент **<figure>**, в элемент **<head>**, либо в один или в несколько элементов **<p>**; это относится и к любым текстам, входящим в состав собственно графического изображения. Настоятельно рекомендуется вводить в элемент **<figDesc>** текстовое описание изображения, такое описание можно будет использовать в приложениях, не воспроизводящих графику, а также для того, чтобы представить документы в форме, доступной для читателей с ослабленным зрением. (Подобный текст обычно не рассматривается как часть документа в узком смысле слова.)

В простейшем случае эти элементы отмечают положение графической информации, как показано в следующем примере:

```
<pb n=412>
<figure></figure>
<pb n=413>
```

(Следует отметить, что нельзя пропускать тег окончания, даже если элемент ничего не содержит.) Более обычны случаи, когда графический объект имеет хотя бы идентифицирующее его название, которое следует закодировать с помощью элемента **<head>**. Кроме того, в большинстве случаев удобно закодировать и краткое описание изображения, как это сделано в следующем примере:

```
<figure>
  <head>Mr Fezziwig's Ball</head>
  <figdesc>A Cruikshank engraving showing Mr Fezziwig leading
    a group of revellers.</figdesc>
</figure>
```

Если существует цифровая версия нужной графической информации, предпочтительнее, очевидно, вставлять именно ее в надлежащее место документа. Графические элементы, например, рисунки, обычно хранятся в объектах (файлах), отличных от файлов, содержащих текст документа, и используют другое представление (формат хранения). TEI Lite DTD поддерживает графику стандартов CGM, TIFF или JPEG, в системе обозначений SGML они называются, соответственно, **cgm**, **tiff** и **jpeg**<sup>1</sup>.

Какой бы формат ни использовался для кодирования изображения, он вводится в документ совершенно одинаково. Первый шаг — это объявление объекта SGML конкретного типа, которое задает название объекта, его внешний идентификатор (например, имя файла) и используемое представление (формат). Например, предположим, что цифровое изображение мяча мистера Феццивига (Mr. Fezziwig's ball) хранится в формате TIFF в файле **fezzi.tff**. Тогда необходимо объявить следующий объект:

```
<!ENTITY fezziPic SYSTEM "fezzi.tff" NDATA tiff>
```

---

<sup>1</sup> Можно использовать, однако, и другие представления, при условии, что в DTD добавлено соответствующее объявление NOTATION; см. главу о таблицах, формулах и графике элементы в TEI P3 или любой справочник по SGML, где приведены подробности об объявлении NOTATION в SGML.

Все подобные объявления должны обрабатываться раньше самого документа SGML; в TEI Lite DTD можно достичь этого, включив их в файл **litedecls.ent** или в любой файл с общим идентификатором **//TEI U5-1995/DTD TEI Lite 1.0 Extensions//EN**.

Если сделано упомянутое выше объявление, для внедрения цифрового изображения в выбранное место документа остается только указать значение для атрибута **entity** элемента **<figure>**:

```
<figure entity==fezziPic>
  <head>Mr Fezziwig's Ball</head>
  <figdesc>A Cruikshank engraving showing Mr Fezziwig leading
    a group of revellers.</figdesc>
</figure>
```

## 16 Интерпретация и анализ

Часто можно услышать, что *любая* разметка — это форма интерпретации или анализа. Хотя четко разграничить "объективную" и "субъективную" информацию каким-нибудь универсальным способом действительно трудно, а может, и невозможно, справедливо также и то, что утверждения, касающиеся последней, обычно считаются более полемичными, чем утверждения, касающиеся первой. Многие школы поэтому предпочитают фиксировать такие интерпретации, только если возможно предупредить читателя о том, что они считаются более спорными (открытыми для обсуждения), чем остальная разметка. В этом разделе описаны некоторые элементы, предусмотренные в системе TEI, с помощью которых можно решить подобные задачи.

### 16.1 Орфография

Интерпретации обычно разбросаны по всему тексту, безотносительно к другим структурным единицам. Следовательно, в качестве первого шага при интенсивном использовании интерпретаций полезно разделить текст на дискретные, идентифицируемые единицы, каждой из которых можно присвоить метку для использования в качестве "канонической ссылки". Чтобы их использование было более легким, эти единицы не должны пересекаться друг с другом или быть вложенными друг в друга. Удобно представлять их с помощью следующего элемента:

**<s>** идентифицирует **s-единицу** (сегмент) в документе, устанавливая простую систему канонических ссылок, которая охватывает весь текст. Предусмотрены следующие атрибуты:

**type** определяет тип единицы (например, **declarative** (повествовательный), **interrogative** (вопросительный) и т.п.)

Как видно из названия, элемент **<s>** в большинстве случаев используется (по крайней мере, в лингвистических приложениях) для маркировки **орфографических предложений**, т.е. единиц, определяемых орфографическими характеристиками, например, пунктуацией. Так, отрывок из романа *Джен Эйр*, ранее уже использованный в примерах, можно разделить на s-единицы следующим образом:

```
<pb n='474'>
<div1 type=chapter n='38'>
<p><s n=001>Reader, I married him.</s>
<s n=002>A quiet wedding we had:</s>
<s n=003>he and I, the parson and clerk, were alone present.</s>
<s n=004>When we got back from church, I went
into the kitchen of the manor-house, where Mary was cooking the dinner,
and John cleaning the knives, and I said &dash;</s>
<p><q><s n=005>Mary, I have been married to Mr Rochester this morning.</s></q> ...
```

Поставленные в этом случае теги концов элементов не являются строго необходимыми, поскольку элементы **<s>** не могут вкладываться друг в друга: начало одного элемента **<s>** предполагает, что предыдущий закончен. Если s-единицы отмечены тегами как показано выше, целесообразно отметить тегами весь текст от начала до конца, так что каждое анализируемое слово будет содержаться только в одном элементе **<s>**, идентификатор которого затем можно использовать для определения уникальной ссылки на него. Если указанные идентификаторы являются уникальными в пределах данного документа, предпочтительнее применять атрибут **id** вместо атрибута **n**, который использован в приведенном выше примере.



## 16.2 Элементы интерпретации общего назначения

Элемент сегментации более общего назначения `<seg>` уже упоминался в настоящем документе как средство для идентифицирования не отмеченных иным способом “целей” перекрестных ссылок и гипертекстовых связей (см. раздел 8); он идентифицирует фрагмент текста на уровне фраз, которому кодировщик может присвоить определенный пользователем **тип**, а также определенный пользователем уникальный идентификатор; этим можно воспользоваться для выделения тегами особенностей текста, которые не предусмотрены в опубликованных “Принципах TEI”.

В “Принципах TEI”, например, не предусмотрен элемент `<apostrophe>` для маркировки тех частей текста, где рассказчик непосредственно адресуется к читателю (или слушателю). Один из способов решения этой проблемы — рассматривать такие фрагменты как случаи использования элемента `<q>`, которые отличаются от других только соответствующим значением атрибута **who**. Возможно, более простым и, безусловно, более общим решением, будет использование элемента `<seg>` следующим образом:

```
<div1 type==chapter n='38'>
<p><seg type='apostrophe'>Reader, I married him.</seg>
A quiet wedding we had: ...
```

Атрибут **type** в элементе `<seg>` может иметь любое значение и, таким образом, может использоваться для регистрации явлений на уровне фраз, относящихся к любым типам; с практической точки зрения полезно фиксировать в заголовке используемые значения и их значимость.

Элемент `<seg>` определенного типа (в отличие от элемента `<s>`, на который он несколько похож) может находиться внутри другого элемента `<seg>` того же или другого типа. Это позволяет создавать весьма сложные структуры; некоторые примеры приведены выше, в разделе 8.3. Однако, поскольку они должны удовлетворять требованию SGML о том, что элементы могут быть вложены друг в друга, но не могут пересекать друг с друга, при этом оказывается невозможным удовлетворить общее требование о связывании интерпретации с произвольными сегментами текста, так как это положение полностью игнорирует иерархию документа. Кроме того, необходимо, чтобы сама интерпретация была представлена единым закодированным значением в атрибуте **type**.

От каких-либо ограничений свободен элемент `<interp>`, который позволяет закодировать сложную интерпретирующую информацию относительно простым способом.

**<interp>** содержит интерпретирующую аннотацию, которую можно связать с некоторой частью текста. Предусмотрены следующие атрибуты:

- value** идентифицирует конкретное аннотируемое явление.
- resp** указывает, кто является ответственным за данную интерпретацию.
- type** указывает, явление какого типа отмечается в данном отрывке. Предусмотрены значения **image** (изображение), **character** (символ), **theme** (тема), **allusion** (аллюзия) или название конкретного типа текста, примеры которого идентифицированы.
- inst** указывает примеры анализа или интерпретации, представленные текущим элементом.

**<interpGrp>** объединяет элементы `<interp>`.

Эти элементы позволяют кодировщику указать и класс интерпретации, и конкретный пример класса, который интерпретируется. Таким образом, если с помощью элемента `<seg>` можно просто указать, что нечто является апострофом (обращением), то с помощью элемента `<interp>` утверждается, что это — пример (апострофа) более широкого класса явлений (риторические фигуры).

Более того, `<interp>` является пустым элементом, который необходимо связать с отрывком, к которому он относится, с помощью или атрибута **ana** (описан выше, в разделе 8.3), или его собственного атрибута **inst**. Это означает, что можно выполнить любой вид анализа, не принимая во внимание иерархию документа SGML, а также облегчить группировку случаев анализа конкретного вида. Последнее достигается с помощью элемента специального назначения `<interpGrp>`.

Например, пусть нужно отметить такие различные аспекты текста, как его тему или предмет, риторические фигуры и места, где происходят отдельные сцены повествования. Различные фрагменты используемого в качестве образца отрывка из *Джен Эйр*, например, можно связать с такими риторическими фигурами, как апострофа, гипербола и метафора; с такими ссылками на тему (предмет), как церкви, слуги, приготовление пищи, почта и медовый месяц; со сценами, происходящими в церкви, на кухне, в неуказанном месте (в гостиной?).

Эти интерпретации можно размещать в любом месте элемента `<text>`; однако, как показывает практика, лучше всего помещать их все в одном месте (например, в отдельном разделе вводной или заключительной части), как это сделано в следующем примере:

```
<back>
<div1 type='Interpretations'
<interp id='fig-apos' resp='LB, MSM'
  type='figure of speech' value='apostrophe'>
<interp id='fig-hyp' resp='LB, MSM'
  type='figure of speech' value='hyperbole'>
<!-- ... -->
<interp id='set-church' resp='LB, MSM'
  type='setting' value='church'>
<!-- ... -->
<interp id='ref-church' resp='LB, MSM'
  type='reference' value='church'>
<interp id='ref-serv' resp='LB, MSM'
  type='reference' value='servants'>
<!-- ... -->
</p></div>
```

Очевидную избыточность этого кодирования можно значительно сократить, если воспользоваться элементом `<interpGrp>`, чтобы сгруппировать все элементы `<interp>`, у которых общие значения атрибутов. Пример такого кодирования приведен ниже:

```
<back>
<div1 type='Interpretations'>
<interpGrp type='figure of speech' resp='LB, MSM'>
<interp id='fig-apos' value='apostrophe'>
<interp id='fig-hyp' value='hyperbole'>
<interp id='fig-meta' value='metaphor'>
<!-- ... -->
</interpGrp>
<interpGrp type='scene-setting' resp='LB, MSM'>
<interp id='set-church' value='church'>
<interp id='set-kitch' value='kitchen'>
<interp id='set-unspec' value='unspecified'>
<!-- ... -->
</interpGrp>
<interpGrp type='reference' resp='LB, MSM'>
<interp id='ref-church' value='church'>
<interp id='ref-serv' value='servants'>
<interp id='ref-cook' value='cooking'>
<!-- ... -->
</interpGrp>
</p></div>
```

Когда эти элементы интерпретации определены, можно двумя способами связать их с частями текста, к которым они относятся, причем можно воспользоваться как одним, так и обоими способами. Атрибутом **ana** можно воспользоваться в любом подходящем элементе:

```
<div1 type=chapter n='38'>
<p id='P38.1' ana='set-church set-kitch'>
<s id=P38.1.1 ana='fig-apos'>Reader, I married him.</s>
```

В этом примере следует обратить внимание на то, что, поскольку в абзаце упомянуто два места действия (в церкви и на кухне), указаны идентификаторы для обоих.

С другой стороны, элементы `<interp>` могут указывать на любые части текста, к которым они относятся, с помощью собственного атрибута **inst**:

```
<interp id='fig-apos' type='figure of speech' resp='LB, MSM'
  value='apostrophe' inst='P38.1.1'>
<!-- ... -->
<interp id='set-church' type='scene-setting' value='church'
```

```

    inst='P38.1' resp='LB, MSM'>
<interp id='set-kitchen' type='scene-setting' value= 'kitchen'
    inst='P38.1' resp='LB, MSM'>
<!-- ... -->

```

Элемент **<interp>** не ограничен каким-либо конкретным типом анализа. Рассмотренный выше пример литературного анализа — это только одна из возможностей, элемент **<interp>** можно также эффективно использовать для лингвистического анализа частей речи. Например, предложение, рассмотренное в разделе 8.3, предполагает лингвистический анализ, который можно представить следующим образом:

```

<interp id=NP1 type=pos value='noun phrase, singular'>
<interp id=VV1 type=pos value='inflected verb, present-tense singular'>
...

```

## 17 Техническая документация

Хотя основное внимание в этом документе уделено использованию системы TEI для кодирования существующих "доэлектронных" документов, такая же процедура может использоваться и для кодирования новых. При подготовке новых документов (таких как этот, например) настоятельно рекомендуется использование SGML: структура документов в этом случае представлена в явном виде, один и тот же электронный текст может использоваться в различных целях — например, чтобы обеспечить как интерактивные гипертекстовые или просмотрные версии, так и хорошо сформатированные типографские версии документа из общего SGML-оригинала.

Для достижения этих целей в TEI Lite включено небольшое число дополнительных элементов как расширение основного списка TEI DTD; упомянутые дополнительные элементы используются для обозначения характерных особенностей технических документов вообще и связанных с SGML документов в частности.

### 17.1 Дополнительные элементы для технических документов

Для обозначения характерных особенностей технических документов можно воспользоваться следующими элементами:

- <eg>** содержит простой краткий пример к некоторой обсуждаемой технической теме, например фрагмент программы или пример кодирования SGML.
- <code>** содержит краткий фрагмент кода на некотором формальном языке (часто на языке программирования).
- <ident>** содержит идентификатор определенного типа, например, имя переменной либо имя элемента или атрибута SGML.
- <gi>** содержит конкретный тип идентификатора: идентификатор SGML или имя элемента.
- <kw>** содержит ключевое слово на некотором формальном языке.
- <formula>** содержит математическую или химическую формулу, которая может быть представлена в некоторой не-SGML системе. Предусмотрены следующие атрибуты:

**notation** указывает систему записи, используемую для представления тела формулы. По умолчанию принимается значение **tex**, означающее, что формула соответствует стандартам системы обработки текстов TeX.

Приведенный ниже пример показывает, как можно использовать эти элементы для кодирования отрывка из вводного курса по программированию на языке ФОРТРАН:

```

<p>It is traditional to introduce a language with a program like the
following:
<eg>
    CHAR*12 GRTG
    GRTG = 'HELLO WORLD'
    PRINT *, GRTG
    END
</eg></p>
<p>This simple example first declares a variable <ident>GRTG</ident>, in

```

the line `<code>CHAR*12 GRTG</kw>`, which identifies `<ident>GRTG</ident>` as consisting of 12 bytes of type `<kw>CHAR</kw>`. To this variable, the value `<mentioned>HELLO WORLD</mentioned>` is then assigned. This is followed by a `<kw>PRINT</kw>` statement and an `<kw>END</kw>` statement.

Обработывая текст, подобный приведенному выше, с помощью форматирующего приложения, можно “указать” этому приложению приемы правильного форматирования (например, сохранить строки оригинала или использовать другой шрифт). Точно так же, использование таких тегов, как `<ident>` и `<kw>`, значительно облегчает построение правильного указателя.

Элемент `<formula>` используется для того, чтобы включить в текст в качестве отдельной единицы математическую или химическую формулу. Поскольку для представления формул, как правило, используется множество самых разнообразных специальных символов и функций, не встречающихся в обычном тексте, возникает необходимость представить основную часть формулы в специализированной системе записи (в специализированной нотации). Используемую нотацию следует определить с помощью атрибута **notation**, как это показано на следующем примере:

```
<formula notation=tex>
  E = mc^2
</formula>
```

Нотация **Tex** для TEI Lite DTD является стандартной; можно использовать и другие нотации, но их необходимо сначала определить с помощью объявления **notation** в DTD.

Внутри элемента `<formula>` допускается почти любая последовательность символов, если она будет обрабатываться процессором, воспринимающим SGML. Данные, не измененные программой синтаксического разбора, проходят для обработки в приложение, которое определяется заданной нотацией. Единственным исключением из этого правила является то, что программа синтаксического разбора *распознает* всякую последовательность символов, напоминающую окончание тега SGML, т.е. знак “меньше” (<), за которым сразу же следует знак “косая черта” (/) и алфавитный символ. Ниже приведен пример последовательности символов, которая окажется причиной того, что программа синтаксического разбора SGML обнаружит ошибки:

```
<formula notation=tex>
  E = mc^-2</a
</formula>
```

К счастью, встретить последовательность символов `</` в большинстве используемых на практике математических нотаций крайне маловероятно. Если же это произойдет, следует воспользоваться особыми средствами, описание которых не входит в задачу данного руководства (более подробная информация приведена в полной версии “Принципов TEI”).

Эта проблема более актуальна, когда темой технического документа является кодирование SGML, а сам документ кодируется в SGML. Очевидно, что в таком документе необходимо четко разграничить разметку SGML, приводимую в качестве примеров, и собственно разметку документа, причем в примерах очень вероятно появление знаков окончания тегов. Наиболее общим решением является выделение тела каждого SGML-примера как содержащего данные, поскольку программа синтаксического разбора не проверяет их соответствие разметке SGML. Это достигается путем выделения примеров в специальную конструкцию SGML, которая называется **отмеченная секция CDATA**. Ниже приведен пример такой конструкции:

```
<p>A list should be encoded as follows:
<eg><! [ CDATA [
  <list>
  <item>First item in the list</item>
  <item>Second item</item>
</list>
]]>
</eg>
The <gi>list</gi> element consists of a series of <gi>item</gi>
elements.
```

Использованный в примере элемент `<list>` не будет рассматриваться как составная часть самого документа, поскольку он введен в пределах отмеченной секции (которая начинается объявлением специальной разметки `<![CDATA [` и заканчивается символами `]]>`).

Следует отметить также использование элемента `<gi>` для выделения тегами ссылок на имена элементов SGML (или **идентификаторы**) в основной части текста.

## 17.2 Генерируемые разделы

Большинство современных систем обработки документов способны автоматически генерировать целые разделы, например, содержание или предметный указатель. В системе TEI Lite предусмотрен элемент для обозначения места, где следует разместить такой сгенерированный раздел.

**<divGen>** указывает место, где появится текстовый раздел, автоматически сгенерированный приложением для обработки текстов. Предусмотрены следующие атрибуты:

**type** указывает тип генерируемого текстового раздела (указатель, содержание и т.п.), который появится в данном месте. Допустимы следующие значения: **index** (указатель, который будет сгенерирован и вставлен в данном месте), **toc** (содержание), **figlist** (список иллюстраций), **tablist** (список таблиц).

Элемент **<divGen>** можно разместить в любом месте, где допускается появление элемента подраздела, как это показано в следующем примере:

```
<front>
<titlePage> ... </titlePage>
<divGen type=toc>
<div type='Preface'><head>Preface</head> ... </div>
</front>
<body> ... </body>
<back>
<div1><head>Appendix</head> ... </div1>
<divGen type=index n='Index'>
</back>
```

В этом примере продемонстрировано также использование атрибута **type** для обозначения различных типов генерируемых разделов: в первом случае это содержание (**toc**), а во втором — указатель.

Если необходимо закодировать уже существующий указатель или содержание (а не сгенерировать их заново), следует воспользоваться элементом **<list>**, который описан в разделе 12.

## 17.3 Генерирование указателя

В то время как автоматическое создание содержания в правильно размеченном тегами документе обычно не является проблемой, создание указателя хорошего качества зачастую требует более тщательной разметки. Может оказаться недостаточным просто составить список всех частей, отмеченных каким-либо определенным тегом, хотя выбор (например) всех случаев появления таких элементов, как **<term>** или **<name>** часто оказывается хорошей отправной точкой для формирования указателя.

В TEI DTD предусмотрен специальный тег **<index>**, которым можно воспользоваться как для обозначения частей документа, которые следует включить в указатель, так и для определения способа его формирования.

**<index>** отмечает место, где предполагается ввести указатель. Предусмотрены следующие атрибуты:

**level1** задает форму ввода элемента указателя основного (первого) уровня.

**level2** задает форму ввода второго уровня, если этот уровень имеется.

**level3** задает форму ввода третьего уровня, если этот уровень имеется.

**level4** задает форму ввода четвертого уровня, если этот уровень имеется.

**index** указывает, к какому указателю (или нескольким указателям) относится данный элемент указателя.

Например, второй абзац этого раздела мог бы выглядеть следующим образом:

```
...
TEI lite also provides a special purpose <gi>index</gi> tag
<index level1='indexing'>
<index level2='index (tag)' level3='use in index generation'>
which may be used ...
```

Элемент `<index>` можно использовать и для того, чтобы зафиксировать некоторые формы пояснений или аналитической информации. Например, при изучении произведений Овидия с целью сравнительного стилистического анализа весьма желательно зарегистрировать ссылки поэта на различных персонажей. В приведенных ниже строках его *Метаморфоз* подобное исследование должно зафиксировать следующие упоминания поэтом Юпитера (**deus**, **se**, и как подлежащее для сказуемого **confiteor** [во флективной форме номер 227]), на Юпитера в образе быка (**imago tauri fallacis** и как подлежащее для сказуемого **teneo**) и т.д.<sup>2</sup>

```
<l n=3.001>iamque deus posita fallacis imagine tauri
<l n=3.002>se confessus erat Dictaeaque rura tenebat
```

Это требование можно было бы удовлетворить с помощью элемента `<note>`, описанного в разделе 7, или элемента `<interp>`, рассмотренного в разделе 16. Здесь показано, как можно достичь поставленной цели с помощью элемента `<index>`.

Предполагается, что конечной целью является генерирование нескольких указателей: одного — для имен богов (он назван **dn**), другого — для ономастических ссылок (он назван **on**), третьего — для местоименных ссылок (он назван **pr**) и т.д. Ниже приведен пример того, как можно решить эту задачу:

```
<l n=3.001>iamque deus posita fallacis imagine tauri
  <index index="dn" level1="Iuppiter" level2="deus">
  <index index="on" level1="Iuppiter (taurus)"
    level2="imago tauri failacis"></l>
<l n=3.002>se confessus erat Dictaeaque rura tenebat
  <index index="pr" level1="Iuppiter" level2="se">
  <index index="v" level1="Iuppiter" level2="confiteor (v227)">
  <index index="mons" level1="Dicte" level2="rura Dictaea">
  <index index="regio" level1="Creta" level2="rura Dictaea">
  <index index="v" level1="Iuppiter (taurus)"
    level2="teneo (v9)"></l>
```

Для каждого приведенного выше элемента `<index>` статья указателя будет создаваться в соответствующем указателе, причем в качестве заглавного слова используется значение атрибута **level1**, в качестве вторичного ключевого слова — значение атрибута **level2**, которое содержит это слово в именительном падеже. Реальная ссылка будет взята из контекста, в котором появляется элемент `<index>`, т.е. в данном случае его содержит идентификатор элемента `<I>`.

## 18 Национальные алфавиты, диакритические знаки и т.п.

Для тех, кто работает со стандартными формами европейских языков, рекомендации по использованию алфавитов в TEI достаточно просты. Для локального использования можно применять алфавиты, которые поддерживает компьютер и программное обеспечение. Если с помощью имеющегося программного обеспечения трудно прямо с клавиатуры ввести специальные символы, можно определить собственную кодировку символов на клавиатуре (например, представить буквы с ударениями путем ввода сразу же за буквой соответствующего знака ударения или воспользоваться специальными последовательностями символов, появление которых в обычном тексте маловероятно, в частности, вводить **аЕ** вместо **ж**). Для преобразования этих “условных обозначений” в нужные символы можно затем воспользоваться функциями глобального поиска и замены. При работе с оригиналами, в которых используется не-латинский алфавит, но существует стандартный способ транслитерации этого алфавита (например, для древнегреческого языка это бета-код Thesaurus LinguGrc), следует применить именно этот способ. Любая используемая транслитерация должна быть обратимой (это исключает удивительно много способов, широко используемых при обычном письме) и, кроме того, предпочтительно, чтобы для нее не требовались специальные лигатуры, связки или диакритические знаки (что исключает значительное число способов транслитерации, удовлетворяющих первому требованию).

Для обмена файлами между системами используются объектные ссылки SGML, чтобы заменить все символы, не попавшие в приведенный ниже список символов, почти никогда не искажаемых при обмене электронными данными:

```
a b c d e f g h i j k l m n o p q r s t u v w x y z
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
```

<sup>2</sup> Этот анализ взят с разрешения авторов из работы Вилларда Мак-Карти и Бартона Райта *Аналитическая ономастика "Метаморфоз" Овидия* (Издание Принстонского университета, в печати). Сделаны некоторые упрощения.

0 1 2 3 4 5 6 7 8 9

" % & ' ( ) \* + , - . / : ; < = > ? (пробел)

Из этого списка исключены следующие символы, которые, к досаде неподготовленных пользователей, часто неправильно передаются через национальные границы или при пересылке по стандартным глобальным сетям. Тем не менее, при переходе от компьютера Macintosh к PC эти символы, скорее всего, не исказятся:

! # \$ [ " ] ^ ' - \_

Чтобы гарантировать адекватную передачу по многопользовательским сетям, необходимо использовать объектные ссылки для всех букв с ударениями и символов расширенного латинского алфавита, всех не-латинских символов, а также для всех символов, которые не предусмотрены на стандартной клавиатуре компьютера.

При желании кодировщик может воспользоваться в TEI-совместимых файлах своими собственными объектными именами SGML, если для этих имен сделаны стандартные объявления объекта SGML, но стандартные имена (хотя и достаточно длинные) имеют то преимущество, что они очевидны; любой знакомый с английским языком пользователь поймет, какой символ кодируется данным объектом, часто даже не прибегая к справочным спискам. Этим качеством не обладают многие другие схемы представления символов с ударениями.

Объектные имена для символов, упомянутых выше как “несохраняемые” и для символов с ударениями некоторых основных западноевропейских языков, приведены ниже. Списки наборов общедоступных объектов и их содержание имеются в любой работе по SGML, указанной в библиографии; приведенные ниже имена относятся к наборам общедоступных объектов ISO, они широко используются и поэтому рекомендуются здесь.

Если необходимый символ не входит в набор общедоступных объектов, можно сгенерировать имя, придерживаясь правил определения имен, используемых в наборе общедоступных объектов ISO, как это описано ниже:

**digraphs** (диграфы) Для формирования объектных имен диграфов после букв, образующих диграф, добавляется последовательность символов **lig**. Если требуются заглавные буквы, следует ввести обе буквы диграфа в верхнем регистре (необходимо помнить, что регистр букв обычно важен в объектных именах). Примеры: **aelig** (ж), **Aelig** (Ж), **szlig** (β).

**diacritics** (диакритические знаки) и **accents** (ударения) Для формирования объектных имен букв с надстрочными знаками, которые встречаются в большинстве западноевропейских языков, после буквы (в верхнем или нижнем регистре), над которой необходимо поставить надстрочный знак, вводится одна из следующих последовательностей символов:

**umlaut** (умляут) для обозначения умляута или трема (две точки над буквой) используются символы **uml**, например: **auml** (д), **Auml** (Д), **euml** (л), **iuml** (п), **ouml** (ц), **Ouml** (Ц), **uuml** (ь), **Uuml** (Ь).

**acute** (знак ударения) для введения знака ударения (чтобы изобразить буквы некоторых алфавитов или ударную гласную) используется последовательность символов **acute**, например: **aacute** (á), **eacute** (é), **Eacute** (É), **iacute** (í), **oacute** (ó), **uacute** (ú).

**grave** (тупое ударение) для введения знака тупого ударения используется последовательность символов **grave**, в частности: **agrave** (ă), **egrave** (ë), **igrave** (î), **ograve** (ò), **ugrave** (Û).

**circumflex** (значок “^” над буквой) для введения этого диакритического знака используется последовательность символов **circ**, например: **acirc** (â), **ecirc** (ê), **Ecirc** (Ê), **icirc** (î), **ocirc** (ô), **ucirc** (û).

**tilde** (тильда) чтобы ввести буквы с тильдами, используется последовательность символов **tilde**, в частности: **atilde** (ã), **Atilde** (Г), **ntilde** (ñ), **Ntilde** (Ñ), **otilde** (õ), **Otilde** (Õ).

**consonants** (согласные буквы). Для некоторых специальных согласных, имеющих в западноевропейских языках, рекомендуются следующие объектные имена: **ccedil** (з), **Ccedil** (З), **eth** (eth нижнего регистра или англо-саксонское/исландское перечеркнутое d), **ETH** (eth верхнего регистра), **thorn** (торн (руническая буква, соответствующая th) нижнего регистра), **THORN** (торн верхнего регистра) **szlig** (немецкая лигатура “s-z” или esszett).

**punctuation marks** (знаки препинания). Для некоторых распространенных знаков препинания рекомендуются следующие объектные имена: **ldquo** (левая двойная кавычка, напоминающая число 66 в виде верхнего индекса), **rdquo** (правая двойная кавычка, напоминающая число 99 в виде верхнего индекса), **mdash** (длинное тире), **hellip** (горизонтальное многоточие, три близко расположенные точки), **rsquo** (одинарная правая кавычка, напоминающая число 9 в виде верхнего индекса). См. также приведенный ниже список “несохраняемых” символов.

"unsafe" characters ("несохраняемые" символы). Символы, перечисленные выше как "несохраняемые" при обмене информацией в современных международных специализированных и общедоступных сетях, можно представить следующими объектами: **excl** (!), **num** (#), **dollar** (\$), **lsqb** (левая квадратная скобка), **bsol** (косая черта с обратным наклоном — \), **rsqb** (правая квадратная скобка), **circ** (значок "^"), **lsquo** (левая одинарная кавычка), **grave** (значок тупого ударения), **lcbub** (левая фигурная скобка, {), **rcub** (правая фигурная скобка, }), **verbar** (вертикальная черта, |), **tilde** (тильда, ~).

## 19 Вводная и закрывающая части

### 19.1 Вводная часть

Материал, предвещающий текст, такой как титульные листы, вступительное слово и т.п., может содержать очень ценную дополнительную лингвистическую или социальную информацию, которая может оказаться полезной для множества различных целей (в особенности это касается старинных текстов). РЗ дает ряд рекомендаций для выделения текстовых элементов, наиболее часто встречающихся во вступительной части, краткое описание которых приводится в данном разделе.

#### 19.1.1 Титульный лист

Начало титульного листа следует отметить элементом **<titlePage>**. Весь размещенный на нем текст следует ввести и отметить с помощью соответствующего элемента из следующего списка:

<b>&lt;titlePage&gt;</b>	содержит титульный лист текста, который появляется во вводной или закрывающей части.
<b>&lt;docTitle&gt;</b>	содержит название документа, включая все его составляющие, в том виде, как они приведены на титульном листе. Должен быть разделен на элементы <b>&lt;titlePart&gt;</b> .
<b>&lt;titlePart&gt;</b>	содержит подразделы или части названия произведения, как представлено на титульном листе; этот элемент используется также для свободно размещаемых фрагментов титульного листа, не являющихся частями названия документа, для указания авторства и т.д. Предусмотрены следующие атрибуты: <b>type</b> определяет роль этих подразделов названия. Допустимы следующие значения: <b>main</b> (основное название), <b>sub</b> (подзаголовок), <b>desc</b> (описание произведения, включенное в название) и <b>alt</b> (альтернативное название).
<b>&lt;byline&gt;</b>	содержит первичные сведения об ответственности, приведенные на титульном листе, либо в начале или в конце произведения.
<b>&lt;docAuthor&gt;</b>	содержит имя автора документа, как приведено на титульном листе (часто, но не всегда содержится в элементе <b>&lt;byline&gt;</b> ).
<b>&lt;docDate&gt;</b>	содержит дату документа, как представлено (обычно) на титульном листе.
<b>&lt;docEdition&gt;</b>	содержит сведения о редактировании, как представлено на титульном листе документа.
<b>&lt;docImprint&gt;</b>	содержит сведения об издании (место и дату публикации, имя издателя), как представлено (обычно) в нижней части титульного листа.
<b>&lt;epigraph&gt;</b>	содержит цитату, анонимную или с указанием автора, которая приводится в начале раздела, главы или на титульном листе.

Где необходимо, следует выделить различные начертания шрифта с помощью атрибута **rend**, как это было описано выше. В "Принципах TEI" пока не предусмотрено подробное описание размещения и размеров букв, используемое в орнаментальных заглавиях. Смену языка следует обозначить с помощью атрибута **lang** или элемента **<foreign>**, в зависимости от того, что необходимо использовать в данном случае. Собственные имена, где бы они не встречались, следует отметить с помощью элемента **<name>**.

Ниже приведены два варианта титульного листа:

```
<titlePage rend=Roman>
  <docTitle><titlePart type=main>
    PARADISE REGAIN'D. A POEM In IV <hi>BOOKS</hi>.
  </titlePart>
  <titlePart>
    To which is added <title>SAMSON AGONISTES</title>.
```



```

    </titlePart>
</docTitle>
<byLine>The Author <docAuthor>JOHN MILTON</docAuthor></byLine>
<docImprint><name>LONDON</name>,
    Printed by <name>J.M.</name>
    for <name>John Starkey</name>
    at the <name>Mitre</name>
    in <name>Fleetstreet</name>,
    near <name>Temple-Bar.</name>
</docImprint>
<docDate>MDCLXXI</docDate>
</titlePage>

<titlePage>
    <docTitle><titlePart type=main>
    Lives of the Queens of England, from the Norman
    Conquest; </titlePart>
    <titlePart type='sub'>with anecdotes of their courts.
    </titlePart></docTitle>
<titlePart>Now first published from Official Records
    and other authentic documents private as well as
    public. </titlePart>
<docEdition>New edition, with corrections and
    additions</docEdition>
<byline>By <docAuthor>Agnes Strickland</docAuthor></byline>
<epigraph>
    <q>The treasures of antiquity laid up in old
        historic rolls, I opened.</q>
    <bibl>BEAUMONT</bibl>
</epigraph>
<docImprint>Philadelphia: Blanchard and Lea</docImprint>
<docDate>1860. </docDate>
</titlePage>

```

### 19.1.2 Вступительная часть

Крупные блоки текста во вводной части следует выделять с помощью элементов `<div>` или `<div1>`. Для того, чтобы отличить друг от друга обычно встречающиеся разные типы вступительных частей текста, используются следующие значения атрибута **type**:

**foreword** текст, адресованный читателю от имени автора, редактора или издателя, (возможно, в форме письма).

**preface** текст, адресованный читателю от имени автора, редактора или издателя, (возможно, в форме письма).

**dedication** текст (часто в виде письма), адресованный не читателю, а определенному лицу, в котором автор обычно просит это лицо обратить внимание на посвящаемый ему труд.

**abstract** текст, кратко передающий содержание произведения.

**ack** Благодарности.

**contents** содержание (обычно оно выделяется с помощью элементов `<list>`).

**frontispiece** иллюстрированный фронтиспис, может включать и некоторый текст.

Как и любой текстовый раздел, разделы вступительной части могут содержать структурные элементы низкого уровня или не-структурные элементы, как уже описывалось ранее. Обычно они начинаются с заголовка (или названия) определенного вида, который следует выделить с помощью элемента `<head>`. Вступительное слово (послание, письмо) может содержать следующие дополнительные элементы:

**<salute>** содержит приветствие, предваряющее предисловие, посвятельное послание или другой раздел текста, а также приветствие в конце письма, предисловия и т.п.

**<signed>** содержит заключительное приветствие и т.п., следующее за предисловием, посвящением или другим разделом текста.

<b>&lt;byline&gt;</b>	содержит первичные сведения об ответственности, представленные на титульном листе, в начале или в конце произведения.
<b>&lt;dateline&gt;</b>	содержит краткое описание места, дату, время и т.п. создания письма, газетной статьи или другого произведения, предваряющее или завершающее его как особый вид заголовка или заключительной части.
<b>&lt;argument&gt;</b>	Формальный список или описание тем, раскрываемых в данном подразделе текста.
<b>&lt;cit&gt;</b>	Цитата из некоторого другого документа, а также библиографическая ссылка на ее источник.
<b>&lt;opener&gt;</b>	группирует вместе элементы <b>&lt;dateline&gt;</b> , <b>&lt;byline&gt;</b> , <b>&lt;salutation&gt;</b> или подобную им информацию, которая появляется в качестве вступительной части в начале раздела, особенно в начале письма.
<b>&lt;closer&gt;</b>	объединяет элементы <b>&lt;dateline&gt;</b> , <b>&lt;byline&gt;</b> , <b>&lt;salutation&gt;</b> или подобную им информацию, которая появляется в качестве закрывающей части в конце раздела, особенно в конце письма.

Послания, которые находятся в любом другом месте текста, будут, конечно, содержать те же самые элементы.

Например, посвящение в начале произведения Джона Мильтона *Comus*, следует разметить следующим образом:

```
<div type='dedication'>
<head>To the Right Honourable <name>JOHN Lord Viscount
BRACLY</name>, Son and Heir apparent to the Earl of
Bridgewater, &amp;c.</head>
<salute>MY LORD,</salute>

<p>THis <hi>Poem</hi>, which receiv'd its first occasion of
Birth from your Self, and others of your Noble Family ....
and as in this representation your attendant
<name>Thyrsis</name>, so now in all reall expression
<closer>
<salute>Your faithfull, and most humble servant</salute>
<signed><name>H. LAWES.</name></signed>
</closer>
</div>
```

## 19.2 Закрывающая часть

### 19.2.1 Структурные подразделения закрывающей части

Из-за различий в издательской практике, закрывающая часть может, в сущности, содержать любой из элементов, перечисленных выше для вводной части, а если так, то и использовать следует те же самые элементы. Кроме того, закрывающая часть может содержать следующие типы сведений в элементе **<back>**. Подобно структурным разделам основного текста, их следует выделять с помощью элементов **<div>** и **<div1>**, а различаются они по следующим допустимым значениям атрибута **type**:

**appendix** приложение

**glossary** список слов и определений, обычно в форме **list type=gloss**.

**notes** ряд элементов **<note>**.

**bibliography** ряд библиографических ссылок, обычно в форме специального элемента для библиографического списка **<listBibl>**, пунктами которого являются отдельные элементы **<bibl>**.

**index** набор элементов указателя, представленный обычно как структурированный список или список-гlossарий, может предваряться (необязательно) элементом **<head>** и, возможно, несколькими абзацами вводного или заключительного текста (в TEI P3 определены и другие специализированные элементы для генерирования указателей при создании документа, они описаны в разделе 17.3).

**colophon** сведения в конце книги, содержащие информацию о том, где, когда и кем она издана; в современных изданиях здесь часто приводят производственные подробности и указывают шрифты, использованные в книге).

## 20 Электронный титульный лист

Каждый текст TEI имеет заголовок, содержащий информацию, аналогичную той, что приводится на титульном листе опубликованного текста. Заголовок вводится с помощью элемента **<teiHeader>** и делится на четыре основные части:

- <fileDesc>** содержит полное библиографическое описание электронного файла
- <encodingDesc>** фиксирует связь между электронным текстом и оригиналом или оригиналами, по которым он составлен.
- <profileDesc>** содержит детальное описание небиблиографических аспектов текста, в частности, используемых в тексте языков и подязыков; ситуации, в которой происходило создание текста; участников и их установок.
- <revisionDesc>** содержит архив внесенных в файл исправлений.

Совокупность или набор текстов, имеющих одни и те же характеристики, может иметь один заголовок для этой совокупности и отдельные заголовки для каждого из ее компонентов. В этом случае атрибут **type** указывает тип заголовка.

```
<teiHeader type=corpus>
```

вводит заголовок для информации на уровне совокупности текстов.

Некоторые элементы заголовка содержат довольно пространные тексты, состоящие из одного или нескольких элементов **<p>**. Другие группируются следующим образом:

- Элементы, имена которых оканчиваются на **Stmt** (для сведений), обычно составляют группу элементов, фиксирующих некоторую структурированную информацию.
- Элементы, имена которых оканчиваются на **Decl** (для объявлений), содержат информацию о применении конкретного кодирования.
- Элементы, имена которых оканчиваются на **Desc** (для описаний) содержат текстовые описания.

### 20.1 Описание файла

Элемент **<fileDesc>** является обязательным. Он содержит полное библиографическое описание файла, которое выполняется с помощью следующих элементов:

- <titleStmt>** группирует информацию о названии произведения и о тех, кто несет ответственность за его содержание.
- <editionStmt>** группирует информацию, относящуюся к одной редакции текста.
- <extent>** описывает приблизительный размер электронного текста, хранящегося на некотором носителе, размер указывается в любых удобных единицах.
- <publicationStmt>** группирует информацию, касающуюся публикации или распространения электронных или других текстов.
- <seriesStmt>** группирует информацию о **серии** (если таковая имеется), к которой принадлежит данная публикация.
- <notesStmt>** объединяет все примечания, содержащие информацию о тексте, которая дополняет ту, что зафиксирована в других частях библиографического описания.
- <sourceDesc>** содержит библиографическое описание экземпляра текста (или текстов), который служил оригиналом для создания или генерирования электронного варианта.

Минимальный заголовок имеет следующую структуру:

```
<teiHeader>
  <fileDesc>
    <titleStmt> ... </titleStmt>
    <publicationStmt> ... <publicationStmt>
    <sourceDesc> ... <sourceDesc>
  </fileDesc>
</teiHeader>
```

### 20.1.1 Сведения о названии

В элементе `<titleStmt>` можно использовать следующие элементы:

- `<title>` содержит название произведения, будь то статья, книга, журнал или серия, включая все альтернативные названия или подзаголовки.
- `<author>` в библиографической ссылке содержит имя автора (авторов) произведения, индивидуального или коллективного; это **первичные сведения об ответственности** для любой библиографической единицы.
- `<sponsor>` указывает название финансирующей организации или учреждения.
- `<funder>` указывает имя лица, название учреждения или организации, ответственных изыскание средств на выполнение проекта или на создание текста.
- `<principal>` содержит имя главного разработчика, ответственного за создание электронного текста.
- `<respStmt>` содержит **сведения об ответственности** за содержание текста, редактирование, запись или серию, если специализированных элементов для авторов, редакторов и т.д. недостаточно или они не применяются.

Рекомендуется, чтобы по названию можно было отличить компьютерный файл от оригинала текста, например:

```
[title of source]: a machine readable transcription
[title of source]: electronic edition
A machine readable version of: [title of source]
```

Элемент `<respStmt>` содержит следующие компоненты:

- `<resp>` содержит информацию, описывающую тип персональной интеллектуальной ответственности за приводимые в тексте сведения.
- `<name>` содержит собственное имя или название.

Пример:

```
<titleStmt>
  <title>Two stories by Edgar Allen Poe: a machine readable
    transcription</title>
  <author>Poe, Edgar Allen (1809-1849)
  <respStmt><resp>compiled by</resp>
  <name>James D. Benson</name></respStmt>
</titleStmt>
```

### 20.1.2 Сведения о редакциях текста

В элементе `<editionStmt>` сгруппирована информация об одной редакции текста (где **edition** используется как везде в библиографии), и этот элемент может включать следующие элементы:

- `<edition>` описывает особенности одной редакции текста.
- `<respStmt>` содержит сведения об ответственности за содержание текста, редактирование, запись или серию, если специализированных элементов для авторов, редакторов и т.д. недостаточно или они не применяются.

Пример:

```
<editionStmt>
  <edition n=U2>Third draft, substantially revised
  <date>1987</date>
</edition>
</editionStmt>
```

Точно определить, что именно составляет новую редакцию электронного текста, должен специалист, кодирующий текст.

### 20.1.3 Сведения о размере

Элемент `<extent>` описывает приблизительный размер файла.

Пример:

```
<extent>4532 bytes</extent>
```

#### 20.1.4 Сведения о публикации

Элемент **<publicationStmt>** является обязательным. Он может содержать простое описание или группу элементов, которые описаны ниже:

**<publisher>** содержит название организации, ответственной за публикацию или распространение библиографической единицы.

**<distributor>** содержит имя лица или название учреждения, ответственного за распространение текста.

**<authority>** содержит имя лица или название учреждения, ответственного за доступность электронного файла (не пугать с издателем или распространителем).

Должен быть представлен хотя бы один из этих трех элементов, если не приводится описание для всей публикации в целом. Эти элементы, в свою очередь, могут содержать следующие элементы:

**<pubPlace>** содержит название места издания библиографической единицы.

**<address>** содержит почтовый или любой другой адрес, например, издателя, организации или отдельного лица

**<idno>** содержит любой стандартный или нестандартный номер, используемый для идентификации библиографической единицы. Предусмотрены следующие атрибуты:

**type** указывает классификацию, которой соответствует номер, например, индекс ISBN или другая стандартная классификация.

**<availability>** содержит информацию о доступности текста, например, любые ограничения на его использование или распространение, авторские права и т.п. Атрибуты включают:

**status** содержит код, определяющий текущую доступность текста. Допустимы следующие значения: **restricted** (ограниченный доступ), **unknown** (неопределенный уровень доступа) и **free** (свободный доступ).

**<date>** содержит дату в любом формате.

Пример:

```
<publicationStmt>
  <publisher>Oxford University Press</publisher>
  <pubPlace>Oxford</pubPlace> <date>1989</date>
  <idno type=ISBN> 0-19-254705-5</idno>
  <availability>Copyright 1989, Oxford University
    Press</availability>
</publicationStmt>
```

#### 20.1.5 Сведения о сериях и примечаниях

Элемент **<seriesStmt>** объединяет информацию о серии (если она существует), к которой относится публикация. Он может содержать элементы **<title>**, **<idno>** или **<respStmt>**.

Элемент **<notesStmt>**, если он используется, состоит из одного или нескольких элементов **<note>**. Некоторая информация, в обычной библиографии находящаяся в примечаниях, в TEI кодируется с помощью специальных элементов.

#### 20.1.6 Описание источников

Обязательный элемент **<sourceDesc>** фиксирует подробности об источнике или источниках, на основании которых создан компьютерный файл. В нем может содержаться простое описание или библиографическая ссылка, в которой используется один или несколько следующих элементов:

**<bibl>** содержит библиографическую ссылку произвольной структуры, компоненты которой можно либо явно отметить тегами, либо не выделять.

**<biblFull>** содержит полностью структурированную библиографическую ссылку, в которой присутствуют все компоненты описания файла TEI.

**<listBibl>** содержит список библиографических ссылок любого вида.

Примеры:

```
<sourceDesc>
  <bibl>The first folio of Shakespeare, prepared by Chariton
    Hinman (The Norton Facsimile, 1968)</bibl>
</sourceDesc>

<sourceDesc>
  <scriptStmt id=CN12>
  <bibl><author>CNN Network News
    <title>News headlines
    <date>12 Jun 1989
  </bibl>
  </scriptStmt>
</sourceDesc>
```

## 20.2 Описание кодирования

Элемент **<encodingDesc>** указывает методы и принципы редактирования, согласно которым записан текст. Настоятельно рекомендуется использовать этот элемент. В нем может содержаться текстовое описание или элементы из следующего списка:

- <projectDesc>** подробно описывает цель создания электронного файла, а также содержит другую необходимую информацию, касающуюся процедуры компоновки или комплектования текста.
- <samplingDecl>** содержит описание причин и методов, используемых для отбора текстов при создании совокупности или собрания текстов.
- <editorialDecl>** содержит подробности о принципах и практике редактирования, которые применялись при кодировании текста.
- <tagsDecl>** содержит подробную информацию о тегах, примененных в документе SGML.
- <refsDecl>** указывает, как построены канонические ссылки в данном тексте.
- <classDecl>** содержит одну или несколько таксономий, определяющих все классификаторные коды, используемые в тексте.

### 20.2.1 Описание проекта и способа отбора текстов

Примеры использования элементов **<projectDesc>** и **<samplingDesc>**:

```
<encodingDesc>
  <projectDesc>Texts collected for use in the Claremont
    Shakespeare Clinic, June 1990.
  </projectDesc>
</encodingDesc>

<encodingDesc>
  <samplingDecl>Samples of 2000 words taken from the beginning
    of the text
  </samplingDecl>
</encodingDesc>
```

### 20.2.2 Определение способов редактирования

Элемент **<editorialDecl>** содержит текстовое описание практических методов, используемых при кодировании текста. Обычно в этом описании отражены следующие темы (каждую из которых удобно вводить в виде отдельного абзаца):

**correction** определяет, как именно и при каких обстоятельствах правился текст.

**normalization** указывает, до какой степени исходный текст регулируется или нормализуется.

**quotation** указывает, что было сделано со знаками кавычек, которые были в оригинале: были ли они оставлены или заменены объектными ссылками, отличаются ли друг от друга открывающие и закрывающие кавычки и т.п.

**hyphenation** указывает, что было сделано с дефисами, которые были в оригинале (в особенности с переносами в конце строк): были ли они оставлены, заменены объектными ссылками и т.п.

**segmentation** указывает, как именно был сегментирован текст, например, по предложениям, по интонационным единицам, по графемам и т.п.

**interpretation** указывает, какая аналитическая или интерпретируемая информация была добавлена в текст.

Пример:

```
<editorialDecl>
  <p>The part of speech analysis applied throughout
    section 4 was added by hand and has not been
    checked.
  <p>Errors in transcription controlled by using the
    WordPerfect spelling checker.
  <p>All words converted to Modern American spelling
    using Webster's 9th Collegiate dictionary.
  <p>All quotation marks converted to entity
    references &odq; and &cdq;.
</editorialDecl>
```

### 20.2.3 Объявление тегов, ссылок и классификаций

Элемент **<tagsDecl>** используется для того, чтобы обеспечить подробную информацию о тегах SGML, которые используются в тексте. Этот элемент может содержать простой список используемых элементов, со счетчиком появлений для каждого из них, для которого используются следующие специальные элементы:

**<tagUsage>** содержит информацию об использовании конкретного элемента в пределах самого внешнего элемента **<text>** документа TEI. Атрибуты включают:

**gi** имя (идентификатор) элемента, отмеченного данным тегом.

**occurs** указывает количество появлений этого элемента в тексте.

Элемент **<rendition>** используется в документе для того, чтобы указать различные способы выделения элементов в исходном тексте.

**<rendition>** содержит информацию о способе, которым следует выделить один или несколько элементов.

**<tagUsage>** содержит информацию об использовании определенного элемента в пределах элемента **<text>**. Предусмотрены следующие атрибуты:

**occurs** указывает число появлений данного элемента в тексте.

**ident** указывает число появлений в тексте данного элемента, который имеет другое значение глобального атрибута **id**.

**render** содержит идентификатор элемента **<rendition>**, который определяет, как именно следует выделить данный элемент.

Например:

```
<tagsDecl>
  <tagUsage gi=text occurs=1>
  <tagUsage gi=body occurs=1>
  <tagUsage gi=p occurs=12>
  <tagUsage gi=hi occurs=6>
</tagsDecl>
```

Приведенное объявление тегов (искусственное) допустимо для текста, содержащего двенадцать абзацев в основной части, причем в ней выделено шесть элементов **<hi>**. Следует отметить, что если используется элемент **<tagsDecl>**, он должен содержать элемент **<tagUsage>** для *каждого* элемента, выделенного тегом в соответствующем элементе текста.

Элемент **<refsDecl>** используется в документе для фиксирования способа построения стандартной системы ссылок в кодируемом тексте. В простейшей форме этот элемент состоит из текстового описания.

Пример:

```
<refsDecl>
  <p>The N attribute on each DIV1 and DIV2 contains the
  canonical reference for each such division in the form
  XX.yyy where XX is the book number in roman numeral and
  yyy is the section number in arabic.
</refsDecl>
```

Элемент **<classDecl>** группирует определения или источники для всех схем классификации, используемых другими частями заголовка. Должна быть определена по крайней мере одна такая схема, закодированная с помощью следующих элементов:

- <taxonomy>** определяет применяемую для классификации текстов типологию либо неявно, посредством библиографических ссылок, либо явно, с помощью структурированной таксономии.
- <bibl>** содержит библиографическую ссылку произвольной структуры, компоненты которой можно либо явно отметить тегами, либо не выделять.
- <category>** содержит отдельную описательную категорию в таксономии, определенной пользователем; эта категория может быть "вложенной" в более общую категорию, т.е. иметь в таксономии любой уровень.
- <catDesc>** описывает в форме краткого текстового описания определенную категорию в пределах таксономии или типологии текста.

В простейшем случае таксономия определяется библиографической ссылкой, как показано в следующем примере:

```
<classDecl>
  <taxonomy id='LCSH'>
    <bibl>Library of Congress Subject Headings
    </bibl>
  </taxonomy>
</classDecl>
```

Специалист, кодирующий электронный текст, может определить альтернативную (или дополнительную) систему классификации для специальных целей, как показано ниже:

```
<taxonomy id=B>
  <bibl>Brown Corpus</bibl>
  <category id==B.A><catDesc>Press Reportage
    <category id=B.A1><catDesc>Daily</category>
    <category id=B.A2><catDesc>Sunday</category>
    <category id=B.A3><catDesc>National</category>
    <category id=B.A4><catDesc>Provincial</category>
    <category id=B.A5><catDesc>Political</category>
    <category id=B.A6><catDesc>Sports</category>
    ...
  </category>
  <category id==B.D><catDesc>Religion
    <category id=B.D1><catDesc>Books</category>
    <category id=B.D2><catDesc>Periodicals and tracts</category>
  </category>
  ...
</taxonomy>
```

Связь между определенным текстом и категорией в такой таксономии определяется с помощью элемента **<catRef>**, который находится внутри элемента **<textClass>**, как будет описано ниже.

## 20.3 Описание параметров

Элемент **<profileDesc>** позволяет регистрировать информацию, характеризующую различные описательные аспекты текста, в одном блоке. Этот элемент имеет три необязательных компонента:

- <creation>** содержит информацию о создании текста.
- <langUsage>** описывает языки, подязыки, наречия, диалекты и т.д., встречающиеся в



тексте.

**<textClass>** группирует информацию, описывающую тип или тему текста в терминах стандартной классификации, тезаурус и т.п.

Примеры:

```
<creation>
  <date value='1992-08'>August 1992</date>
  <name type=place>Taos, New Mexico</name>
</creation>
```

Элемент **<textClass>** классифицирует текст, ссылаясь на систему или системы, определенные в элементе **<classDecl>**; может содержать один или несколько следующих элементов:

**<keywords>** содержит список ключевых слов или фраз, идентифицирующих тему или тип текста. Предусмотрены следующие атрибуты:

**scheme** указывает управляемый словарь, в котором определен набор соответствующих ключевых слов.

**<classCode>** содержит использованный для данного текста код классификации по некоторой стандартной классификации. Предусмотрены следующие атрибуты:

**scheme** идентифицирует используемую систему классификации или таксономию.

**<catRef>** указывает одну или несколько определенных категорий в пределах некоторой таксономии или типологии текста. Предусмотрены следующие атрибуты:

**target** идентифицирует соответствующие категории.

Элемент **<keywords>** содержит список ключевых слов или фраз, идентифицирующих тему или тип текста. Атрибут **scheme** связывает их с системой классификации, определенной в элементе **<taxonomy>**.

```
<textClass>
  <keywords scheme=LCSH>
    <list>
      <item>English literature -- History and criticism --
        Data processing.</item>
      <item>English literature -- History and criticism --
        Theory etc. </item>
      <item>English language -- Style - Data
        processing.</item>
    </list>
  </keywords>
</textClass>
```

## 20.4 Описание исправлений

Элемент **<revisionDesc>** обеспечивает протокол изменений, в котором регистрируется каждое изменение текста. Протокол может быть составлен как последовательность элементов **<change>**, каждый из которых содержит:

**<date>** содержит дату в любом формате.

**<respStmt>** содержит сведения об ответственности за содержание текста, редактирование, запись или серию, если специализированных элементов для авторов, редакторов и т.д. недостаточно или они не применяются.

**<item>** содержит один компонент списка.

Пример:

```
<revisionDesc>
  <change><date>6/3/91:</date>
    <respStmt><name>EMB</name><resp>ed.</resp></respStmt>
    <item>File format updated</item>
  <change><date>5/25/90:</date>
    <respSmt><name>EMB</name><resp>ed.</resp>
    <item>Stuart's corrections entered</item>
</revisionDesc>
```

## 21 Список рассмотренных элементов

### 21.1 Глобальные атрибуты

- Все элементы для определения типа TEI Lite-документов имеют следующие глобальные атрибуты:
- ana** связывает элемент и его интерпретацию.
  - corresp** связывает элемент с одним или несколькими соответствующими элементами.
  - id** уникальный идентификатор элемента; должен начинаться с буквы, может содержать буквы, цифры, дефисы и точки.
  - lang** язык текста в данном элементе; если значение не указано, считается, что в элементе использован тот же язык, что и в окружающем контексте.
  - n** имя или номер этого элемента; может использоваться любая последовательность символов. Часто применяется для записи традиционных систем ссылок.
  - next** связывает данный элемент со следующим элементом в совокупности.
  - prev** соединяет данный элемент с предыдущим элементом в совокупности.
  - rend** указывает способ начертания элемента в оригинале текста: **italic** (курсив), **roman** (прямой светлый шрифт), **display block** (выделенный блок) и т.п. Значением может быть любая строка символов.

### 21.2 Элементы TEI Lite

В следующем списке перечислены все элементы, определенные в TEI Lite DTD, и приведено их краткое описание:

- |                             |   |
|-----------------------------|---|
| <b>&lt;abbr&gt;</b>         | содержит сокращение любого типа; полная форма может быть задана атрибутом <b>expan</b> .  |
| <b>&lt;add&gt;</b>          | содержит буквы, слова или фразы, введенные в текст автором текста, вступительного слова, аннотации или корректором.   |
| <b>&lt;address&gt;</b>      | содержит почтовый или любой другой адрес, например, издателя, организации или отдельного лица.  |
| <b>&lt;addrLine&gt;</b>     | содержит одну строку почтового или любого другого адреса.   |
| <b>&lt;anchor&gt;</b>       | определяет место или точку в документе, чтобы на них можно было сослаться.  |
| <b>&lt;argument&gt;</b>     | Формальный список или описание тем, раскрываемых в данном подразделе текста.  |
| <b>&lt;author&gt;</b>       | в библиографической ссылке содержит имя автора (авторов) произведения, персональное или имя коллективного автора; это первичные <b>сведения об ответственности</b> для любой библиографической единицы. |
| <b>&lt;authority&gt;</b>    | содержит имя лица или название учреждения, ответственного за доступность электронного файла (не путать с издателем или распространителем).  |
| <b>&lt;availability&gt;</b> | содержит информацию о доступности текста, например, любые ограничения на его использование или распространение, авторские права и т.п.  |
| <b>&lt;back&gt;</b>         | содержит различные приложения и т.п., которые располагаются после основной части текста.  |
| <b>&lt;bibl&gt;</b>         | содержит библиографическую ссылку произвольной структуры, компоненты которой можно либо явно отметить тегами, либо не выделять.   |
| <b>&lt;biblFull&gt;</b>     | содержит полностью структурированную библиографическую ссылку, в которой присутствуют все компоненты описания файла TEI.  |
| <b>&lt;biblScope&gt;</b>    | определяет степень подробности (“объем”) библиографической ссылки, например, указаны ли в ней только номера страниц или приведены также названия подразделов большого произведения.                     |
| <b>&lt;body&gt;</b>         | содержит всю основную часть одного монолитного текста, исключая то,   |

	что относится к вводящей или закрывающей частям текста.
<byline>	содержит первичные сведения об ответственности, приведенные на титульном листе, либо в начале или в конце произведения.
<catDesc>	описывает в форме краткого текстового описания определенную категорию в пределах таксономии или типологии текста.
<category>	содержит отдельную описательную категорию в таксономии, определенной пользователем; эта категория может быть "вложенной" в более общую категорию, т.е. иметь в таксономии любой уровень.
<catRef>	указывает одну или несколько определенных категорий в пределах некоторой таксономии или типологии текста.
<cell>	содержит одну ячейку таблицы.
<cit>	Цитата из некоторого другого документа, а также библиографическая ссылка на ее источник.
<classCode>	содержит использованный для данного текста код классификации по некоторой стандартной классификации, которая определяется атрибутом <b>scheme</b> .
<classDecl>	содержит одну или несколько таксономий, определяющих все классификаторные коды, используемые в тексте.
<closer>	объединяет элементы <dateline>, <byline>, <salutation> или подобную им информацию, которая появляется в качестве закрывающей части в конце раздела, особенно в конце письма.
<code>	содержит краткий фрагмент кода на некотором формальном языке (часто на языке программирования).
<corr>	содержит правильную форму отрывка, который в тексте приведен с явными ошибками.
<creation>	содержит информацию о создании текста.
<date>	содержит дату в любом формате, нормализованное значение которой хранится в атрибуте <b>value</b> .
<dateline>	содержит краткое описание места, дату, время и т.п. создания письма, газетной статьи или другого произведения, предваряющее или завершающее его как особый вид заголовка или заключительной части.
<del>	содержит букву, слово или отрывок, удаленные, отмеченные как удаленные, либо другим способом обозначенные как излишние или неверные в рукописном тексте автором, переписчиком, составителем аннотации или корректором.
<distributor>	содержит имя лица или название учреждения, ответственного за распространение текста.
<div>	содержит раздел вводного, основного или закрывающего текста.
<div1>...<div7>	содержит разделы первого, второго... седьмого уровня текста вводной, основной или закрывающей части.
<divGen>	указывает место, где появится текстовый раздел, автоматически сгенерированный приложением для обработки текстов; атрибут <b>type</b> указывает, является ли этот раздел указателем, содержанием или имеет какой-либо еще тип.
<docAuthor>	содержит имя автора документа, как приведено на титульном листе (часто, но не всегда содержится в элементе <byline>).
<docDate>	содержит дату документа, как представлено (обычно) на титульном листе.
<docEdition>	содержит сведения о редактировании, как представлено на титульном листе документа.
<docImprint>	содержит сведения об издании (место и дату публикации, имя издателя), как представлено (обычно) в нижней части титульного листа.
<docTitle>	содержит название документа, включая все его составляющие, в том виде, как они приведены на титульном листе. Должен быть разделен на

	элементы <b>&lt;titlePart&gt;</b> .
<b>&lt;edition&gt;</b>	описывает особенности одной редакции текста.
<b>&lt;editionStmt&gt;</b>	группирует информацию, относящуюся к одной редакции текста.
<b>&lt;editor&gt;</b>	вторичные <b>сведения об ответственности</b> для библиографической единицы, например, имя конкретного лица либо название учреждения или организации (возможно, нескольких), выступающих в роли редактора, составителя, переводчика и т.п.
<b>&lt;editorialDecl&gt;</b>	содержит подробности о принципах и практике редактирования, которые применялись при кодировании текста.
<b>&lt;eg&gt;</b>	содержит простой краткий пример к некоторой обсуждаемой технической теме, например фрагмент программы или пример кодирования SGML.
<b>&lt;emph&gt;</b>	отмечает слова или фразы, выделяемые с целью получения лингвистического или риторического эффекта.
<b>&lt;encodingDesc&gt;</b>	фиксирует связь между электронным текстом и оригиналом или оригиналами, по которым он составлен.
<b>&lt;epigraph&gt;</b>	содержит цитату, анонимную или с указанием автора, которая приводится в начале раздела, главы или на титульном листе.
<b>&lt;extent&gt;</b>	описывает приблизительный размер электронного текста, хранящегося на некотором носителе, размер указывается в любых удобных единицах.
<b>&lt;figure&gt;</b>	отмечает место, где в документ будет вставлена графика. Можно использовать атрибуты, чтобы указать объект SGML, содержащий само изображение (в некоторой не-SGML нотации); абзацы в элементе <b>&lt;figure&gt;</b> можно использовать для ввода заголовков.
<b>&lt;fileDesc&gt;</b>	содержит полное библиографическое описание электронного файла.
<b>&lt;foreign&gt;</b>	указывает слово или фразу, написанные не на том же языке, что весь остальной текст.
<b>&lt;formula&gt;</b>	содержит математическую или химическую формулу, которая может быть представлена в некоторой не-SGML системе записи. Атрибут <b>notation</b> используется для того, чтобы указать название не-SGML нотации, использованной для записи формулы.
<b>&lt;front&gt;</b>	содержит различную вступительную информацию (заголовки, титульный лист, предисловия, посвящения и т.п.), которую размещают перед основным текстом.
<b>&lt;funder&gt;</b>	указывает имя лица, название учреждения или организации, ответственных изыскание средств на выполнение проекта или на создание текста.
<b>&lt;gap&gt;</b>	указывает точку, где при переписке были пропущены материалы, либо по редакторским причинам, описанным в заголовке TEI, как часть принятой практики выбора, либо из-за того, что эти материалы неудобочитаемы или невняты.
<b>&lt;gi&gt;</b>	содержит конкретный тип идентификатора: идентификатор SGML или имя элемента.
<b>&lt;gloss&gt;</b>	отмечает слово или фразу, которые представляют собой толкование (глоссу) или определение какого-либо другого слова или фразы.
<b>&lt;group&gt;</b>	содержит число монолитных текстов или групп текстов.
<b>&lt;head&gt;</b>	содержит любые заголовки, например, название раздела, заголовков списка или глоссария.
<b>&lt;hi&gt;</b>	отмечает слово или фразу, которые графически отличаются от остального текста, причем причина выделения не указывается.
<b>&lt;ident&gt;</b>	содержит идентификатор определенного типа, например, имя переменной либо имя элемента или атрибута SGML.
<b>&lt;idno&gt;</b>	содержит любой стандартный или нестандартный номер, используемый для идентификации библиографической единицы; атрибут <b>type</b> указывает используемую классификацию или стандарт.

<imprint>	содержит информацию, связанную с публикацией или распространением библиографической единицы.
<index>	отмечает место, где предполагается ввести указатель. Предусмотрены атрибуты для определения основной формы элементов данного указателя, а также определения форм элементов второго - четвертого уровней.
<interp>	содержит интерпретирующую аннотацию, которую можно связать с некоторой частью текста. Предусмотрены следующие атрибуты: <b>resp</b> , <b>type</b> и <b>value</b> .
<interpGrp>	объединяет элементы <interp>.
<item>	содержит один компонент списка.
<keywords>	содержит список ключевых слов или фраз, идентифицирующих тему или тип текста; если ключевые слова выбираются из управляемого словаря, их можно идентифицировать с помощью атрибута <b>scheme</b> .
<kw>	содержит ключевое слово на некотором формальном языке.
<l>	содержит одну (возможно, неполную) стихотворную строку.
<label>	содержит метку, связанную с компонентом списка; в глоссариях отмечает определяемый термин.
<langUsage>	описывает языки, подязыки, наречия, диалекты и т.д., встречающиеся в тексте.
<lb>	отмечает начало новой (типографской) строки в определенном издании или редакции текста.
<lg>	содержит группу стихотворных строк, составляющих формальную единицу, например, строфу, рефрен, стихотворный абзац и т.д.
<list>	содержит любую последовательность единиц, составляющих список, независимо от типа форматирования списка: нумерованные пункты, отмеченные жирной точкой или другим способом.
<listBibl>	содержит список библиографических ссылок любого вида.
<mentioned>	отмечает подразумеваемые, но не используемые слова или фразы.
<milestone>	отмечает границы между разделами текста в соответствии со стандартной системой разметки. Предусмотрены атрибуты <b>ed</b> (редакция), <b>unit</b> (страница или другой тип раздела) и <b>n</b> (новое значение).
<name>	содержит имя собственное, которое может представлять собой фразу. С помощью атрибутов можно указать тип имени, привести его нормализованную форму, или связать его посредством уникального идентификатора с конкретной личностью или предметом.
<note>	содержит примечание или пояснение (аннотацию); с помощью атрибутов можно указать тип, местонахождение в тексте и источник примечания.
<notesStmt>	объединяет все примечания, содержащие информацию о тексте, которая дополняет ту, что зафиксирована в других частях библиографического описания.
<num>	содержит записанное в любой форме число, нормализованное значение которого содержится в атрибуте <b>value</b> .
<opener>	группирует вместе элементы <dateline>, <byline>, <salutation> или подобную им информацию, которая появляется в качестве вступительной части в начале раздела, особенно в начале письма.
<orig>	содержит исходное прочтение, а исправленный вариант может быть приведен в виде значения атрибута <b>reg</b> .
<p>	отмечает абзацы написанного прозой текста.
<pb>	отмечает границу между двумя страницами текста в стандартной системе разметки.
<principal>	содержит имя главного разработчика, ответственного за создание электронного текста.

<profileDesc>	содержит детальное описание небибблиографических аспектов текста, в частности, используемых в тексте языков и подязыков; ситуации, в которой происходило создание текста; участников и их установок.
<projectDesc>	подробно описывает цель создания электронного файла, а также содержит другую необходимую информацию, касающуюся процедуры компоновки или комплектования текста.
<ptr>	ссылка на другое место текущего документа, выраженная одним или несколькими идентифицируемыми элементами.
<publicationStmt>	группирует информацию, касающуюся публикации или распространения электронных или других текстов.
<publisher>	содержит название организации, ответственной за публикацию или распространение библиографической единицы.
<pubPlace>	содержит название места издания библиографической единицы.
<q>	содержит цитату или кажущуюся цитату.
<ref>	ссылка на другое место текущего документа, выраженная одним или несколькими идентифицируемыми элементами, причем в эти элементы можно ввести дополнительный текст или комментарий.
<refsDecl>	указывает, как построены канонические ссылки в данном тексте.
<reg>	содержит прочтение, исправленное или нормализованное по определенным признакам; исходное прочтение можно задать с помощью атрибута <b>orig</b> .
<rendition>	содержит информацию о способе, которым следует выделить один или несколько элементов.
<resp>	содержит информацию, описывающую тип персональной интеллектуальной ответственности за приводимые в тексте сведения.
<respStmt>	содержит сведения об ответственности за содержание текста, редактирование, запись или серию, если специализированных элементов для авторов, редакторов и т.д. недостаточно или они не применяются.
<revisionDesc>	содержит архив внесенных в файл исправлений.
<row>	содержит одну строку таблицы.
<rs>	содержит имя общего назначения или называющую строку. С помощью атрибутов можно указать тип, привести нормализованную форму или связать его с конкретным лицом или предметом посредством уникального идентификатора.
<s>	идентифицирует <b>s-единицу</b> (сегмент) в документе, устанавливая простую систему канонических ссылок, которая охватывает весь текст.
<salute>	содержит приветствие, предваряющее предисловие, посвятельное послание или другой раздел текста, а также приветствие в конце письма, предисловия и т.п.
<samplingDecl>	содержит описание причин и методов, используемых для отбора текстов при создании совокупности или собрания текстов.
<seg>	идентифицирует отрывок или сегмент текста в документе, чтобы на них можно было сослаться; тип сегмента по принятой классификации задается атрибутом <b>type</b> .
<series>	содержит информацию о серии, в которой выходит данная книга или другая библиографическая единица.
<seriesStmt>	группирует информацию о <b>серии</b> (если таковая имеется), к которой принадлежит данная публикация.
<sic>	содержит текст, который следует воспроизвести неизменным, несмотря на его явную некорректность или неточность.
<signed>	содержит заключительное приветствие и т.п., следующее за предисловием, посвящением или другим разделом текста.
<soCalled>	содержит слово или фразу, за которые автор или рассказчик снимает с себя ответственность, что обозначается кавычками или курсивом.

<b>&lt;sourceDesc&gt;</b>	содержит библиографическое описание экземпляра текста (или текстов), который служил оригиналом для создания или генерирования электронного варианта.
<b>&lt;sp&gt;</b>	содержит прямую речь персонажа в тексте пьесы (сценария) или отрывок, представляющий таковую в тексте прозаического или стихотворного произведения; для идентификации говорящего используется атрибут <b>who</b> .
<b>&lt;speaker&gt;</b>	содержит специальную форму заголовка или метки, где указано имя одного или нескольких персонажей, произносящих реплики в пьесе или в некотором фрагменте текста.
<b>&lt;sponsor&gt;</b>	указывает название финансирующей организации или учреждения.
<b>&lt;stage&gt;</b>	содержит различные ремарки к тексту пьесы или к отрывку пьесы.
<b>&lt;table&gt;</b>	содержит текст, выводимый в виде таблицы, т.е. состоящий из строк и столбцов.
<b>&lt;tagsDecl&gt;</b>	содержит подробную информацию о тегах, примененных в документе SGML.
<b>&lt;tagUsage&gt;</b>	содержит информацию об использовании конкретного элемента в пределах самого внешнего элемента <b>&lt;text&gt;</b> документа TEI.
<b>&lt;taxonomy&gt;</b>	определяет применяемую для классификации текстов типологию либо неявно, посредством библиографических ссылок, либо явно, с помощью структурированной таксономии.
<b>&lt;term&gt;</b>	содержит обозначение из слов (одного или нескольких) или символов, которое рассматривается как технический термин.
<b>&lt;textClass&gt;</b>	группирует информацию, описывающую тип или тему текста в терминах стандартной классификации, тезаурус и т.п.
<b>&lt;time&gt;</b>	содержит фразу, определяющую время суток в любом формате; нормализованное значение содержится в атрибуте <b>value</b> .
<b>&lt;title&gt;</b>	содержит название произведения, будь то статья, книга, журнал или серия, включая все альтернативные названия или подзаголовки.
<b>&lt;titlePage&gt;</b>	содержит титульный лист текста, который появляется во вводной или закрывающей части.
<b>&lt;titlePart&gt;</b>	содержит подразделы или части названия произведения, как представлено на титульном листе; этот элемент используется также для свободно размещаемых фрагментов титульного листа, не являющихся частями названия документа, для указания авторства и т.д.
<b>&lt;titleStmnt&gt;</b>	группирует информацию о названии произведения и о тех, кто несет ответственность за его содержание.
<b>&lt;trailer&gt;</b>	содержит заключающее название или “окончание”, которое располагается после текста раздела.
<b>&lt;unclear&gt;</b>	содержит слово, фразу или отрывок, которые нельзя расшифровать с уверенностью из-за неудобочитаемости или невнятности оригинала.
<b>&lt;xptr&gt;</b>	определяет указатель (ссылку) на другое место текущего документа или другого (внешнего) документа.
<b>&lt;xref&gt;</b>	определяет указатель (ссылку) на другое место текущего документа или другого (внешнего) документа; возможно, с дополнительным текстом или комментарием.

## 22 Библиография

В этом приложении содержится список библиографических ссылок на работы по SGML и связанной с ним тематике, представленный в форме, которая демонстрирует использование элемента **<bibl>**, описанного выше в разделе 13.

```

<listBibl>
<bibl>ALA (American Library Association). <title>ALA-LC
Romanization Tables: Transliteration Schemes for Non-Roman
Scripts</title>, approved by the Library of Congress and the American
Library Association, tables compiled and edited by Randall K. Barry.
Washington: Library of Congress, 1991.
</bibl>

<bibl>ANSI (American National Standards Institute). <title>ANSI
X3.4-1986. American National Standard for Information Systems --- Coded
Character Sets --- 7-bit American National Standard Code for Information
Interchange (7-bit ASCII).</title> [New York]: ANSI, 1986.
</bibl>

<bibl>
<author>Barnard, David, et al.</author>
<title level=a>SGML-Based Markup for Literary Texts.</title>
<title>Computers and the Humanities</title>
<biblScope>22 (1988): 265-76.</biblScope>
</bibl>

<bibl>
<author>Barron, David</author>
<title level=a>Why use SGML?</title>
<title>Electronic Publishing
Origination, Dissemination and Design</title>
<biblScope>2.1 (April 1989): 3-24.</biblScope>
</bibl>

<bibl>
<author>Coombs, James H., Alien H. Renear, and Steven J.
DeRose.</author> <title level=a>Markup Systems and the Future of
Scholarly Text Processing. </title> <title>Communications of the
ACM</title>
<biblScope>30.11 (November 1987) : 933-947.</biblScope>
</bibl>

<bibl>
<editor>Cover, Robin C., et al.</editor>
<title>A Bibliography on Structured Text:
Technical Report 90-281</title>
<publisher>Queen's University,</publisher>
<pubPlace>Kingston, Ont.</pubPlace>
<date>June 1990</date>
<note place=inline>A current version of this bibliography
is maintained at <code>http://www.sil.org/sgml/sgml.html</code>.
</bibl>

<bibl>Goldfarb, Charles F., <title>The SGML Handbook.</title>
Oxford: Clarendon Press, 1990. </bibl>

<bibl>
<author>van Herwijnen, Eric.</author>
<title>Practical SGML.</title>
<publisher>Kluwer Academic Publishers</publisher>
<date>1990; 2d ed. 1994.</date>
</bibl>

<bibl>ISO (International Organization for Standardization).
<title>ISO 8859-1: 1987 (E). Information processing --- 8-bit
Single-Byte Coded Graphic Character Sets --- Part 1: Latin Alphabet No.
1.</title> (<title>Traitement de l'information --- Jeux de caracte''res
graphiques cod&eacute;s sur un seui octet --- Partie 1: Alphabet latin no
1.</title>) First edition --- 1987-02-15. [Geneva]: International

```



Organization for Standardization, 1987.

</bibl>

<bibl>ISO (International Organization for Standardization).  
<title>ISO 8879-1986 (E). Information processing --- Text and Office  
Systems --- Standard Generalized Markup Language (SGML).</title> First  
edition --- 1986-10-15. [Geneva]: International Organization for  
Standardization, 1986.

</bibl>

<bibl>ISO (International Organization for Standardization).  
<title>ISO 8879:1986 / A1:1988 (E). Information processing --- Text and  
Office Systems --- Standard Generalized Markup Language (SGML),  
Amendment 1.</title> Published 1988-07-01.

[Geneva]: International Organization for Standardization, 1988.

</bibl>

<bibl>ISO (International Organization for Standardization).  
<title>ISO/TR 9573-1988 (E). Information processing---SGML support  
facilities---Techniques for using SGML.</title> Final text of  
1988-09-12.

</bibl>

<bibl>ISO (International Organization for Standardization), and IEC  
(International Electrotechnical Commission). <title>ISO/IEC 10646-1:  
1993. Information technology --- Universal Multiple-Octet Coded  
Character Set (UCS) --- Part 1: Architecture and Basic Multilingual  
Plane.</title>

[Geneva]: International Organization for  
Standardization, 1993.

</bibl>

<bibl>ISO (International Organization for Standardization), and IEC  
(international Electrotechnical Commission).  
<title>ISO/IEC 10744: 1992. Information  
Technology --- Hypermedia/Time-based Structuring Language  
(HyTime).</title>

[Geneva]: International Organization for Standardization, 1992.

</bibl>

<bibl>

Langendoen, D. Terence, and Gary F. Simons.

<title level=a>A Rationale for the TEI  
Recommendations for Feature-Structure Markup.</title>

<title>Computers and the Humanities</title>

(1995; in press).

</bibl>

<bibl>

<author>Warmer, J., and S. van Egmond</author>

<title level=a>The implementation of the Amsterdam  
SGML parser.</title>

<title>Electronic Publishing

Origination, Dissemination and Design</title>

<biblScope>2.2 (July 1989): 65-90.</biblScope>