Manual to the SimpléTEXML package: 1. Structure and basic macrodefinitions

V. A. Nechitailenko[1]

Received 10 January 2012; accepted 25 January 2012; published 26 January 2012.

[1] The format of modern scientific article is not limited to three hundred years tradition of "text-graphics-table" and assumes the inclusion of modern tools for displaying scientific data and analysis results, and interaction with scientific publications online, as well as compatibility with the modern systems of indexing and dissemination of scientific content. The technologies of electronic publishing developed and used by the Geophysical Center RAS aim to solve at least some of mentioned problems. We consider here the new style file SimpléTEXML which belongs to the family of packages that extend the standard LATEX-class `article.cls` and supports the basic elements of the semantics of articles required for the "on-the-fly" generation of the XML meta-description compliant with the CrossRef XML schema, ver. 4.3.0. The package is developed for journals, published by the GC RAS and registered in the CrossRef system, but can easily be extended to other journals. This manual is mainly addressed to authors who use the LATEX system to prepare their articles for publication. *KEYWORDS: academic journal, electronic publications, EP technologies, information technology, record of science, including semantic, meta-description of articles; CrossRef; online interactive presentations; Geophysical Center RAS.*

Citation: Nechitailenko, V. A. (2011), Manual to the SimpléTEXML package: 1. Structure and basic macrodefinitions, *Russ. J. Earth. Sci., 3,* ES1002, doi:10.2205/2011NZ000101.

Introduction

[2] The SimpléTEXML package is the latest in a series of LATEX2ε packages developed for the journals and books published by the Geophysical Center RAS. Some parts of this manual is geared more to the authors, rather than the publisher's technical staff. The author hopes that the paper will be useful for those who are familiar with the LATEX2ε publishing system and prefer to use it for typesetting their articles.

[3] NB! This paper is not a substitute for or alternative to the well-known guides to LATEX2ε publishing system. It contains information about the specific features of the style of scientific articles published in journals of GC RAS, and describes a number of redefined macros of `article` class, as well as a number of new macros defined in the SimpléTEXML package, which simplify essentially marking up source tests. For more detailed information on the LATEX2ε publishing system and numerous extensions one should refer to the well-known works [*Goossens et al.*, 1993; *Lamport*, 1994; *Mittelbach et al.*, 2006, etc.].

[4] The most significant feature of the package is implemented in its possibility to generate an XML file containing the meta-description of the article compliant with the XML schema of CrossRef, ver. 4.3.0.

The Package Structure and Basic Options

The package structure

[5] The current version of the package extends and modifies the LATEX2ε class `article.cls` and presented in the form of LATEX2ε style file `simpletexml.sty`. Additional style files are called directly from the package except `hyperref` package, which is called from a preamble of the translated article. This is done for the convenience of switching on/off some options of `hyperref` package, e.g. turn off the generation of PDF contents while working with Russian texts.[2] The package contains the following components:

[6] **Generation of meta-description of the article.** To do this, there are five macros that include a set of operations to create and open the file `\jobname.xml` with writing to this file a number of parameters extracted

[1]Geophysical Center RAS, Moscow, Russia

http://elpub.wdcb.ru/journals/rjes/doi/2011NZ000101.html

[2]Adobe Reader version, available at the time of this writing, does not have Russian language support at the system level.

from the input data stream in the process of translation of the document (article). In general, this process takes place "behind the screen", file `\jobname.xml` is created along with other supplementary files (`\jobname.pdf`, `\jobname.aux`, `\jobname.log` , etc.).

[7] **Hyperlinks to internal and external objects** are based on a macro definitions of `hyperref` package. The references to internal objects are also defined "behind the screen" that allows easier typing and avoid errors. For links to the DOI indices special macros based on `hyperref` package are used.

[8] **Supported journal styles.** The SimpleTEX*XML* package supports styles of the following journals:

- International Journal of Geomagnetism and Aeronomy
- Russian Journal of Earth Sciences
- Geophysical Papers Online – Preprints
- Russian Journal of Earth Sciences (in Russian)
- Vestnik Otdelenia nauk o Zemle RAN (in Russian)

The basic package options

[9] The SimpleTEX*XML* package defines options which can be specified in the preamble.

- `\pdfout{dvi|pdf}` – generation of `.dvi` or `.pdf` files;
- `\xml@out{xml|noxml}` – generation of `.xml` file during the translation process;
- `\warnmes{verbose|silent}` – termination of translation process with sending error/warning messages on the screen, or sending them to `.log` file only;
- `\@lang{eng|rus}` – document language (English or Russian);
- `\parnumber{nopar|par}` – paragraphs without or with numbers;
- `\hyptest{hyper|nohyp}` – active or non-active hyperlinks.

[10] The first values are default ones. To customize different values, they should be included into the list of package options. For example, this manual was translated with the options `silent, pdf, rus, par`, i.e. `\usepackage[silent, pdf,rus,par]{simpletexml}`. The number and order of customizable options are arbitrary.

The Structure of the Article

[11] The article prepared for translation must have the following structure:

```
\documentclass[twoside]{article}
\usepackage[<custom_options>]{simpletexml}
\usepackage[<custom_options>]{hyperref}
<text_of_preamble>
\begin{document}
<title/author[s]/organisation[s]>
\abstract{<abstract_text>}
<structured_article_text>
\acknowledgement[s]{<acknowledgment_text>}
                (may be skipped)
\begin{references}
<list_of_cited_publications>
\end{references}
\noreferences   - used instead of
        \begin/end{references} if list is empty
\appendix      (may be skipped)
\supplement     - used to attach electronic
                    supplements.
\end{document}
```

Preamble

[12] The preamble of the article should include the following minimal set of parameters:

```
\journalid{<journal_code>}{<volume>}{<number>}
          {<publication_date>}
\cpright{<copyright_owner>}{<copyright_year>}
\paperid{<CCC_code,_if_applicable>}
        {<article_number_in_issue>}
\papercode{<DOI_suffix>}
\lefthead{<running_head_(author)>}
\righthead{<running_head_(short_title)>}
\received{<day>}{<month>}{<year>}
[\revised{<day>}{<month>}{<year>}]
\accepted{<day>}{<month>}{<year>}
\published{<>}{<month>}{<year>}
\keywords{<keywords_vs_author's_choice>}
\indexterms{<keywords_vs_standard_list>}
\authaddr{<author[s]_address[es]>}
```

[13] The parameter `<journal_code>` is the key parameter, which is used to select predefined values for the journal, i.e. fonts for various elements of the article, the style of the title page, the set of corresponding metadata, ISSN, DOI prefix, etc.

[14] The following journal codes are predefined:

`GI` International Journal of Geomagnetism and Aeronomy
`ES` Russian Journal of Earth Sciences (RJES)
`RE` Russian Journal of Earth Sciences (in Russian)
`NZ` Vestnik Otdelenia nauk o Zemle RAN (in Russian)
`PO` Geophysical Papers Online – Preprints
`MS` Manuscript (unformatted version of the article).

[15] The parameter `<DOI_suffix>` is a string of 12 characters. The suffix may be arbitrary. Usually it is composed according to the rules set by the publisher. For GC RAS

journals the first four characters – the year of acceptance of an article for publication, the fifth and sixth characters – journal code (capital letters), the last six characters – the registration number of an article. For example, DOI suffix of the original version of this article, published in Rusian in January 2011 is `2011NZ000101`.

[16] The running head is composed of two parameters `\lefthead` and `\righthead`, defined in the preamble. Parameters `\received`, `\revised`, `\accepted` and `\published` are used to generate the imprint data.

[17] The `\keywords` parameter includes words or phrases on the choice of the author which reflect in the best way the theme of the article. The `\indexterms` parameter includes a set of keywords from a predefined list, for example, AGU keywords, GCMD keywords, etc.

[18] The parameter `\authoraddr` contains the address(es) of authors in the following format: A. First, B. Second, First Organization, City, State/Region, Country (email address). If authors present more than one organization it is necessary to add additional instructions `\authoraddr`, whose number must match the number of organizations.

[19] Some additional parameters, such as `\subident`, `\xmldepositor`, `\xmlregistrant` are also included in the preamble (details are given in the SimpleTEXML package).

Title page

[20] The main elements of the cover page are inserted after the command `\begin{document}`. Among them `\title{<article_title>}`, set of commands describing the author(s), organization(s) and their internal links and the abstract content (`\abstract{<abstract_content>}`).

[21] A description of each author includes seven parameters, of which 1–3 and 5 are mandatory, others may be absent.

```
\author{<author_number>}{<last_name>}{<given_names_
or_initials>}{[<suffix>]}{<organisation_number>}
{[<alternative_organisation>]}{[<initials>]}
```

For example,

```
\author{1}{Bender}{Ostap-Suleyman Berta-Maria}
{}{1}{}{O.-S. B.-M.},
\author{2}{Allen}{Joe H.}{}{2}{}{J. H.},
```

or

```
\author{1}{Bender}{O.-S. B.-M.}{}{1}{}{},
\author{2}{Allen}{J. H.}{}{2}{}{},
```

Here `<suffix>` – name suffix, e.g., `Jr., Esc.` and the like.

[22] The description of each organization consists of five parameters, all of them except the fourth one, are mandatory, for example,

```
\affil{<organisation_number>}{<organisation>}
{<city>}{<[region/state]>}{<country>}
```

[23] Below is an example of authors' block description[3]

```
\author{1}{First}{A. B.}{}{1}{}{}
\author{2}{Second}{C. D.}{}{2}{}{}
\author{3}{Third}{E. F.}{}{3}{4}{}
\author{4}{Fourth}{G. H.}{}{5}{6}{}
\author{5}{Fifth}{I. J.}{}{1}{}{}
\affil{1}{Institute or Irreproducible Physics}
    {Bigcity}{}{Gallardia}
\affil{2}{Institute of Cosmetic Physics}{Smalltown}
    {}{Gallardia}
\affil{3}{Center of Paranormal Studies}{Miraclecity}
    {}{Gallardia}
\affil{4}{Also at Virtual Affiliation}{Moon City}{}
    {Gallardia}
\affil{5}{Unknown Affiliation}{Somewhere}{}
    {Gallardia}
\affil{6}{Also at Institute of Cosmetic Physics}
    {Smalltown}{}{Gallardia}
```

[24] The number of authors, which is used by template for putting commas correctly, is defined during the first translation and saved in the `.aux` file. A warning message appears on the screen and in the `.log` file with the request to repeat translation. Double translation is also necessary if the number of authors in the editing process was changed, though in this case the warning is not generated.

[25] Numbers of reference links to affiliations should be in ascending mode, i.e. 1, 2, 3, 4, 5, but not 1, 3, 2, 4. Repeating of reference already used is possible, e.g. 1, 2, 2, 3, 4, 1. See also the sample above.

[26] Basic and alternative addresses should be placed in the same order as in the authors' string. This is checked during the translation and warning is issued.

[27] Addresses of alternative affiliations should start with `Also at, Formerly at, On leave from`, and the like, as an author wishes.

[28] Please look at the sample above. If the same affiliation is mentioned as a basic one (for the author `Second`) and as an alternative one (for the author `Fourth`) it should appear twice in the list of affiliations, the second time starting with `Also at`.

[3] Some complexity of the structure of the block is caused by the necessity to transfer the set of semantic elements and quantitative characteristics (number of authors, organizations, etc.) to the TEXengine. The TEX engine does not allow to provide character-wise analysis, moreover, the system does not work with arrays. In the process of interpreting the input stream, we need, on the one hand, to ensure independent processing of input (semantic) elements, without which it is impossible to construct a valid `jobname.xml` file, on the other – formatted grammatically correct strings (with the correct punctuation) before these strings will be converted to an output stream. This problem can be solved by the two-step translation. At the first step the basic syntactic structure is determined by the analysis of input parameters, which are written then into `jobname.aux` file, at the second step, with the use of TEX registers, semantically and syntactically correct line of authors, organizations, string citation lists of referred literature, etc. are formed.

[29] The `\abstract{<abstract_text>}` command is the mandatory component of the title page. At least the empty `\abstract{}` command should appear if the article has no abstract. It is necessary because this command starts the process of converting input to output and the formation of the title page. At the same time all other page elements are generated "behind the screen" using the data defined in the preamble.

Floats

[30] Floating objects (*floats* in LaTeX2ε terms) are elements of articles, not having a fixed point in the text. Place of their appearance is determined by the program in the translation process in accordance with the algorithm implemented in the used class or package. The main float environments in the current version of SimpléTEX*ML* are figures, tables, and panels.

[31] In-text references to `Figure 1` and `Figures 1 and 2` should be marked up as `\figref{1}` and `\figrefs{1 and 2)` or `\figrefs{1, 2}`.

[32] In-text references to `Table 1` and `Tables 1 and 2` should be marked up as `\tabref{1}` and `\tabrefs{1 and 2)` or `\tabrefs{1, 2}`.

[33] In-text references to `Plate 1` and `Plates 1 and 2` should be marked up as `\plaref{1}` and `\plarefs{1 and 2)` or `\plarefs{1, 2}`.

[34] The value in curly brackets for the expressions of the type `Figure 2a`, `Table 3b`, use only the digital part number, i.e. `\figref{2}a` and `\tabref{3}b`, since it is the numeric part of numbers used by the appropriate macro to generate pairs "`anchor-target`".

[35] Graphic files should be named in the order they are mentioned in the text, for example, `f01`, `f02`, etc. The full name of the graphic file can be arbitrary, but it is recommended to use `\paperid` command of the preamble as the prefix in graphics name. That is, the full name of the graphic file in the description, for example, the second figure of the article should be `\setimage{<x-bias>}{<y-bias>}{<width>}{<height>}{2010nz000062-f02}` for `\paperid{2010NZ000062}`. In addition to the name of the graphic file is required at least one of the options or `<width>` or `<height>`.

[36] The sample of figure description is given below:

```
\begin{figure*}[t]                          % Fig 5
\figurewidth{35pc}
\setimage{}{}{}{33pc}{2010nz000062-f05}
\caption{Areas of daily driven fire from June to
August 2010 in the European part of Russia~(a) and
the Moscow region~(b) (``Aerocosmos'' data).}
\end{figure*}
```

Hyperlinks

In-text hyperlinks

[37] Hyperlinks inserted in the text must be verified not only for the correctness of their syntax, but also on the efficiency, i.e. they should be checked for compliance DOI or URL address of the calling document. The main command for hyperlinks – `\href{<URL_address>} {<anchor_text>}`, for example, `\href{http://onznews.wdcb.ru}{Vestnik ONZ RAN}`.

[38] **Sample 1.**

Initial text fragment, e.g.

```
... Dr. So-and-so published in Vestnik ONZ RAN the
article explaining...
```

After converting the initial text to LaTeX2ε format we have

```
...Dr. So-and-so published in \href{http://onznews.
wdcb.ru}{Vestnik ONZ RAN} the article explaining...,
```

This fragment after translation by `\pdflatex` looks like

"... Dr. So-and-so published in Vestnik ONZ RAN the article explaining..."

with a working hyperlink from the PDF file to a specified address.[4]

[39] **Sample 2.**

Initial text fragment, e.g.

```
...AIRS data of the AQUA satellite
[http://mirador.gsfc.nasa.gov] ...
```

should be converted to

```
... AIRS data of the AQUA satellite [\href
{http://mirador.gsfc.nasa.gov}{http://mirador.gsfc.
nasa.gov}] ...
```

If both arguments of the command `\href{}{}` are the same it can be replaced with the command with one argument, i.e. `\url{<target_address>}`

[40] Both fragments, namely,

```
...AIRS data of the AQUA satellite [\href{http://
mirador.gsfc.nasa.gov}{http://mirador.gsfc.nasa.
gov}]...
```

and

```
...AIRS data of the AQUA satellite[\url{http://
mirador.gsfc.nasa.gov}]...
```

after translation by `\pdflatex`, look identical, i.e.

"...AIRS data of the AQUA satellite [`http://mirador.gsfc.nasa.gov`]..."

[41] In case when the string `<target_address>` is too

[4] Colored frames of internal and external hyperlinks are displayed when viewing this PDF file on the computer screen. It is suppressed while printing.

long and translator can not break it correctly the command `\href{}{}` become more preferred. In this case the first argument (URL address or DOI index) remains unchanged, which is necessary to guarantee working hyperlink, while a forced break may be included into the second argument (`<anchor_text>`), the latter appears after translation in the PDF text and will be seen as, "...AIRS data of the AQUA satellite (see http://mirador.gsfc.nasa.gov/cgi-bin/mirador/presentNavigation.pl?tree=project&project=AIRS&CGISESSID=cafeec90b5ebf0b0574ad3505419ea62)..."

DOI-based hyperlinks

[42] Following the CrossRef recommendations DOI-based ***in-text*** hyperlink is composed of two parts: DOI Resolver's address (`http://dx.doi.org/`) and DOI index, i.e. for the DOI index `10.2205/2010NZ000062` corresponding URL will be `http://dx.doi.org/10.2205/2010NZ000062`.

[43] Hyperlinks in the ***reference*** section to external objects with assigned DOIs are composed using a few special commands (see next section).

The List of References

[44] ***In-text*** reference to the item in the List of References, for example,

```
...as shown in the work [Mitkin, 2000] effect of...
```

should be converted to

```
...as shown in the work [{\it Mitkin,}
\reflink{mitk00}{2000}] effect of...,
```

which after translation by `\pdflatex` looks like

"...as shown in the work [*Mitkin,* 2000] effect...".

[45] Another ***in-text*** reference format

```
Mitkin  [2000] shown that...
```

should be converted to

```
{\it Mitkin}  [\reflink{mitk00}{2000}] shown...,
```

which after translation by `\pdflatex` looks like

"*Mitkin* [2000] shown... "

[46] Items in the List of References are marked up as shown below, i.e. in the format

```
<author[s]> (<publication_year>), <article_title>,
<journal>, <volume>, <[issue]> <[extra_parameters:
_publisher,_city]>, <first_page>[--<last_page>]
```

or

```
<author[s]> (<publicat._year>), [<article_title>,]
<book|collection_of_articles>, <[vol.]>, <[issue]>,
<publisher>, <city>, <first_page>[--<last_page>]
```

Non-mandatory parameters are inside square brackets.

[47] Once the list of references has been prepared (see the example the *List of References* marked up for this article in Table 1.), You must select it fragments used to construct the meta-description of articles on the translation stage. To do this, use the list of reference markups:

- `\refauth{<last_name_of_first_author>}`
- `\reftitle{<article_title>}`
- `\refyear{<publication_year>}`
- `\refjour{<journal_title>}`
- `\refbook{<book|collection_title>}`
- `\refvol{<volume>}`
- `\refnumb{<issue>}`
- `\refpage{<first_page>}`

[48] If reference item includes DOI index it will be enough to markup only DOI index, as shown below.

Marking up hyperlinks to publications with DOI indices

- `\nbdoi{<DOI_index>}` – for indices with no math symbols like (<, >), and characters having special meaning in the TEX publishing system. The reference [*Mitkin et al.,* 2000] shows this case. The string in the marked up text looks like `doi:\nbdoi{10.1111/j.1751-908X.2000.tb00774.x}`.

- `\brdoi{<DOI_first_fragment>}{<DOI_second_fragment>}` – for indices with no math symbols and special characters, if they are too long and translator can't break them correctly. The reference [*Plyusnina et al.,* 2000] shows this case. The string in the marked up text looks like `doi:\brdoi{10.1016/S0883-}{2927(99)00092-X}`.

- `\tagdoi{<first_fragment>}{<second_fragment>}{<third_fragment>}` – for indices with math and special symbols. In this case unchanged DOI index of the cited document is used as the `<first_fragment>` of the command, while the text used as the anchor is put to the `<second_fragment>` and to the `<third_fragment>` if necessary to break target string. Symbols "`<`" and "`>`" in `<anchor_text>` should be replaced with `$<$` and `$>$` correspondingly, or with `\textless` and `\textgreater`. The reference [*Hildebrand,* 1991] shows this case. The string in the marked up text looks like `\tagdoi{10.1130/0091-7613(1991)019<0867:CCAPCT>2.3.CO;2}{10.1130/0091-7613(1991)019\textless 0867:CCAPCT\textgreater 2.3.CO;2}{}`.

Acknowledgment

[49] If necessary an acknowledgment(s) can be put here using the `\acknowledgment[s]{<acknowledgment_text>}` command.

Table 1. Marked up list of references in this article

```
\reference{lamp94}
\refauth{Lamport} L. (\refyear{1994}), {\itshape \refbook{A Document Preparation System \LaTeX}},
Addison-Wesley, \refpage{272}~pp.

\reference{goos93}
\refauth{Goossens} M., F. Mittelbach, A. Samarin (\refyear{1993}), {\itshape \refbook{The \LaTeX
Companion}}, Addison-Wesley, \refpage{530}~pp.

\reference{mitt06}
\refauth{Mittelbach} F. at al. (\refyear{2006}), {\itshape \refbook{The \LaTeX
Companion}}, Second Edition, Addison-Wesley, \refpage{554}~pp.

\reference{mitk00}
\refauth{Mitkin}, B. N., A. A. Galizky, T. M. Korda (\refyear{2000}),
\reftitle{Some observations on the determination of gold and the platinum-group elements in black shales},
{\it \refjour{Geostandards Newsletter}, \refvol{24}}, \refpage{227}--240,
doi:\nbdoi{10.1111/j.1751-908X.2000.tb00774.x}.

\reference{plyu00}
\refauth{Plyusnina}, L. P., T. V. Kuzmina, G. G. Likhoidov, G.~A.~Narnov
(\refyear{2000}), \reftitle{ Experimental modeling of platinum sorption on organic matter},
{\it \refjour{Applied Geochemistry}, \refvol{15}},
\refpage{777}--784, doi:\brdoi{10.1016/S0883-}{2927(99)00092-X}.

\reference{hild91}
\refauth{Hildebrand} A. R., Penfield G. T., Kring D. A. (\refyear{1991}), \reftitle{Chicxulub
crater: A possible Cretaceous-Tertiary boundary impact crater on the Yucatan peninsula, Mexico},
{\it \refjour{Geology}, \refvol{19},} \refpage{867},
doi:\tagdoi{10.1130/0091-7613(1991)019<0867:CCAPCT>2.3.CO;2}
{10.1130/0091-7613(1991)019\textless0867:CCAPCT\textgreater2.3.CO;2}{}.
```

References

Lamport L. (1994), *A Document Preparation System LaTeX*, Addison-Wesley, 272 pp.

Goossens M., F. Mittelbach, A. Samarin (1993), *The LaTeX Companion*, Addison-Wesley, 530 pp.

Mittelbach F. at al., (2006), *The LaTeX Companion*, Second Edition, Addison-Wesley, 554 pp.

Mitkin, B. N., A. A. Galizky, T. M. Korda (2000), Some observations on the determination of gold and the platinum-group elements in black shales, *Geostandarts Newsletter, 24*, 227–240, doi:10.1111/j.1751-908X.2000.tb00774.x.

Plyusnina, L. P., T. V. Kuzmina, G. G. Likhoidov, G. A. Narnov (2000), Experimental modeling of platinum sorption on organic matter, *Applied Geochemistry, 15*, 777–784, doi:10.1016/S0883-2927(99)00092-X.

Hildebrand A. R., Penfield G. T., Kring D. A. (1991), Chicxulub crater: A possible Cretaceous-Tertiary boundary impact crater on the Yucatan peninsula, Mexico, *Geology, 19,* 867, doi:10.1130/0091-7613(1991)019<0867:CCAPCT>2.3.CO;2.

Nechitailenko V. A., Geophysical Center RAS, Moscow, Russia (vitaly@wdcb.ru).