MlBibTeX: a Survey

Jean-Michel HUFFLEN*

October 2005

Abstract

This short article introduces end-users to the broad outlines of the first publicly available version of MlBibTeX (1.3). In particular, we explain what MlBibTeX is for BibTeX users. Finally, the organisation of the whole of MlBibTeX's documentation is described.
Keywords LaTeX, BibTeX, MlBibTeX, bibliography files, bibliography styles, bst, nbst, XML, XSLT.

Sommario

Questo breve articolo introduce l'utente finale alla prima versione pubblicamente disponibile di MlBibTeX (1.3). In particolare vengono esposte le principali differenze tra BibTeX e MlBibTeX e viene data descrizione dell'intera documentazione di MlBibTeX.
Parole chiave LaTeX, BibTeX, MlBibTeX, bibliografia, stile bibliografico, bst, nbst, XML, XSLT.

Contents

*LIFC (FRE CNRS 2661). University of Franche-Comté. 16, route de Gray. 25030 BESANÇON CEDEX, FRANCE. E-mail: hufflen@lifc.univ-fcomte.fr.

0 Introduction

This article is a short introduction to a new bibliography processor usable to generate 'References' sections for documents written using the LaTeX word processor [22]. It is written using the Scheme programming language[1] and allows an easy connection with XML's[2] world, since it internally uses XML-like representation. Bibliography styles, which control the layout of output bibliographies, are written using a new language, called nbst, for '**N**ew **B**ibliography **ST**yles', close to the XSLT[3] language. This nbst language makes easier the specification of multilingual features. In fact, we developed this bibliography processor, MlBibTeX, in order to ease the implementation of multilinguism in bibliographies. MlBibTeX, for '**M**ulti**L**ingual BibTeX', is a re-implementation of BibTeX [25], the bibliography processor usually associated with LaTeX.

This article is not a reference manual for MlBibTeX, it just surveys this program. Reading this article requires only basic knowledge about LaTeX, BibTeX, and XSLT. It was originally designed to be the companion of the slides presented at the GUIT[4] conference on Saturday 9th October 2004, we have updated it in order to incorporate recent references.

1 A simple example

Let us look at Figure 1. We can recognise BibTeX-like syntax for a **bibliographical entry**, included in a bibliography (.bib) file, and we can notice some extensions of BibTeX's original syntax. Person names can be expressed using more user-friendly syntax, emphasising the four parts of a name—*First*, *von*, *Last*, *Junior*—as they have been defined in BibTeX [25, § 4]. There is also a way to specify the abbreviation of a first name, when it should not be abbreviated using the default way: retaining only the first letter of each component of the first name. Concerning multilingual information, the `LANGUAGE` field specifies the language of an entry[5]. A notation such as '`[...] ! italian`' means that the string enclosed by square brackets is to be put only in a bibliography for an Italian work, the *references*[6] of this bibliography being expressed in Italian as far as possible, according to a *document-dependent approach* [10, § 4]. All these features and approaches are described in detail in [10].

Now let us suppose that MlBibTeX processes this entry because it is cited within a LaTeX document. Parsing this entry results in an XML tree[7], as shown in Figure 2. By the way, let us notice that in Figure 1, we typed accented letters either directly (e.g., '`à`') or by means of LaTeX commands (e.g., '`\'{o}`'). But

[1] More precisely, the current revision of this language, described in [19].

[2] e**X**tensible **M**arkup **L**anguage. Readers interested in an introduction to this metalanguage can refer to [26].

[3] e**X**tensible **S**tylesheet **L**anguage **T**ransformations [31]. This is the language of transformations for XML documents. Readers interested in a didactive introduction to this language can refer to [29].

[4] ***G**ruppo **U**tilizzatori **I**taliani di TEX* (Italian TeX user group).

[5] The value associated with this field does not have to be surrounded by braces or double quotes. See [17] for more details.

[6] W.r.t. MlBibTeX's terminology, an *entry* is an element of a bibliographical database—like the entry `robeson1965b` in Figure 1—; formatting entries result in *references*.

[7] More precisely, in the representation of this XML tree in Scheme using SXML (Scheme implementation of XML) [20].

```
@BOOK{robeson1965b,
        AUTHOR = {first => Kenneth, last => Robeson},
        TITLE = {The Lost Oasis},
        PUBLISHER = {Bantam},
        EDITION = 2,
        SERIES = {Doc Savage Series},
        NUMBER = 6,
        NOTE = {[No es coneix la traducci\'{o} al Català] ! catalan
                [Titre de la traduction française :
                 ‘‘L’oasis perdue’’] ! french
                [Titel der deutschen \"{U}bersetzung:
                 ‘‘Oase der Verlorenen’’] ! german
                [Titolo della traduzione in italiano:
                 ‘‘L’oasi dei diamanti’’] ! italian
                [T\'{i}tulo de la traducci\'{o}n al Espa\~{n}ol:
                 ‘‘1.000.000 de recompensa’’] ! spanish}
        YEAR = 1965,
        MONTH = apr,
        LANGUAGE = english}
```

Figure 1: Example of an entry using MlBibTeX's features.

these accented letters are single characters in Figure 2, since this representation, like any XML text, uses Unicode [28]. Concerning bibliography styles, they can can use XPath[8] expressions to address parts of this tree like in XSLT. For example, putting person names down can be done by the template given in Figure 3: family names are displayed using small capitals, first names are displayed *in extenso* if they are given within the .bib file, like in the 'References' section of the present article. Let us notice the path expressions—`first`, `von`, ...—giving access to the parts of a person name. A slightly more complicated path is given in the named template `no-lastname`.

As shown in Figure 3—a part of a bibliography style that could be applied to the tree given in Figure 2—the nbst language is close to XSLT. The main difference between them consists of using the `language` attribute. To illustrate that by a simple example, let us assume that we are building a bibliography for a work in Italian. If a last name does not exist when we are processing the items of this bibliography, we can define a template that displays an error message in Italian:

```
<nbst:template name="no-lastname" language="italian">
  <nbst:warning>
    Manca il cognome per <nbst:value-of select="../../../@id"/>
  </nbst:warning>
</nbst:template>
```

Such a template, with a `language` attribute, has higher priority than a template without. In fact, the template named `no-lastname`, given in Figure 3, is a *default template*, called when there is no template with the `language` attribute associated with the current language. Of course, this feature is used for other

[8] XPath is a language used to address parts of an XML document [30].

```
<book id="robeson1965b" language="english">
  <author>     <!--  An author element is a list of names, connected by     -->
    <name>     <!--  'and' or 'with'.     -->
      <personname><first>Kenneth</first><last>Robeson</last></personname>
    </name>
  </author>
  <title>The Lost Oasis</title>
  <publisher>Bantam</publisher>
  <year>1965</year>
  <month><apr/></month>
  <number>6</number>
  <series>Doc Savage Series</series>
  <edition>2</edition>
  <note>
    <group language="catalan">No es coneix la traducció al Català</group>
    <group language="french">
      Titre de la traduction française :
      <emph emf="no" quotedbf="yes">L'oasis perdue</emph>
    </group>
    <group language="german">
      Titel der deutschen Übersetzung:
      <emph emf="no" quotedbf="yes">Oase der Verlorenen</emph>
    </group>
    <group language="italian">
      Titolo della traduzione in italiano:
      <emph emf="no" quotedbf="yes">L'oasi dei diamanti</emph>
    </group>
    <group language="spanish">
      Título de la traducción al Español:
      <emph emf="no" quotedbf="yes">1.000.000 de recompensa</emph>
    </group>
  </note>
</book>
```

Figure 2: The entry in Figure 1 viewed as an XML tree.

purposes than just displaying error messages, it allows us to refine a general framework for a particular language. Some examples are given in [15].

2 Using MlBibTeX

Once MlBibTeX is installed, you can use it by:

```
mlbibtex job-name
```

where '*job-name*' is the name—with or without suffix—of an auxiliary (.aux) file[9]. You can force the use of the language of the bibliography of a document by:

```
mlbibtex job-name --language=...
```

[9] See [23, § 12.1.3] for more details.

```
<nbst:template match="personname">
  <nbst:if test="first">
    <nbst:value-of select="first"/>
    <nbst:text> </nbst:text>     <!--  Putting a space character.     -->
  </nbst:if>
  <nbst:if test="von">
    <nbst:value-of select="von"/>
    <nbst:text> </nbst:text>
  </nbst:if>
  <nbst:choose>
    <nbst:when test="last">
      <nbst:text>\textsc{</nbst:text>   <!--  LaTeX will use small    -->
      <nbst:value-of select="last"/>    <!--  capitals for the last   -->
      <nbst:text>}</nbst:text>          <!--  name.     -->
    </nbst:when>
    <nbst:otherwise>
      <nbst:call-template name="no-lastname"/>
    </nbst:otherwise>
  </nbst:choose>
  <nbst:if test="junior">, Jr.</nbst:if>
</nbst:template>

<nbst:template name="no-lastname">
  <nbst:warning>
    There is no last name for <nbst:value-of select="../../../@id"/>
  </nbst:warning>
</nbst:template>
```

Figure 3: Example of templates written using the nbst language.

but we do not recommend this feature: multilingual functions may be used improperly, and some parts of the resulting text can be processed incorrectly by LaTeX.

As explained in [17], MlBibTeX may need to parse a part of the LaTeX source file[10]. If this source file has to be located, because it cannot be deduced from the name of the .aux file, use the `tex-file` option:

```
mlbibtex job-name --tex-file=...
```

3 MlBibTeX for BibTeX users

Bibliography files usable with BibTeX are reusable with MlBibTeX, except that square brackets are now syntactic delimiters, whereas they are 'ordinary' characters in 'old' BibTeX. The second point about the compatibility of such files is that cross-references, given by `CROSSREF` fields [23, § 13.2.5], are not handled by MlBibTeX presently, processing cross-references is planned for the next version. Except for these two points, .bib files suitable for BibTeX should be accepted by MlBibTeX. Some entries may be processed without problem, whereas BibTeX

[10] BibTeX never does this operation.

issued a warning message[11].

About bibliography styles, we designed a compatibility mode, so that 'old' .bst files[12] can be interpreted by MlBibTeX [8], but this feature has not been implemented yet in the current version. Anyway, it is preferable to use bibliography styles written in nbst, in order to take advantage of MlBibTeX's expressive power. Most of the 'plain' and 'alpha' styles (references labelled by numbers or keys like '[Rob65b]')—that is, the styles implementing a number-only system [23, § 12.2]—have been rewritten. That is not done yet for the styles implementing the other systems: author-date, author-number, short-title [23, §§ 12.3–12.6].

4 More information

In this introductory paper, we have given only some broad outlines. Our purpose is to show that MlBibTeX is usable for most end-users of BibTeX. Of course, there is a 'to do' list for future versions, but we also expect some background from users in order to improve it. At the end of October 2005, the program will be downloadable from MlBibTeX's Web page:

`http:lifc.univ-fcomte.fr/~hufflen/texts/mlbibtex/mlbibtex`

where the slides about MlBibTeX—in English, French, and German—shown at various conferences can be found. Most of the articles about this program are available over there, too. Hereafter, we aim to guide users who wish more information and we explain how the items of MlBibTeX's documentation are organised.

- If you wish to get familiar with MlBibTeX, here are six essential references:
  - [10] is a general and detailed presentation, followed by an abridged reference manual of the nbst language and our functions associated with path expressions;
  - [9] explains why we decided to get rid of adapting the bst language and how we begin to design the broad outlines of nbst;
  - [14] shows how to develop an interface between LaTeX and MlBibTeX for a particular language;
  - [15] discusses a method for developing bibliography styles using nbst;
  - [16] contains details about MlBibTeX's installation and the modules of this program; a second part is to appear;
  - [17] explains how the information about natural languages is managed in MlBibTeX.
- The following paper addresses MlBibTeX's features from a point of view related to the implementation:

[11] An example is an entry, being of type `@ARTICLE`, `@BOOK`, or `@BOOKLET` with both `AUTHOR` and `EDITOR` fields. From our personal point of view, such a choice allows editors of a work to be specified separately from the complete list of its authors. An example is given by the current report on Scheme[19]: depending on the bibliography style used, the complete list is printed *in extenso*—as we do at the end of this article—or skipped—as we did in the bibliography of [18]—; of course, editors' names are put down in both cases.

[12] That is, bibliography styles written using the language described in [24].

  - [18] describes the organisation of the implementation in Scheme.

- The following references are of historical interest:
  - [3] is the first article about MlBibTeX, in French, sketching its first version;
  - [4] is the reference of the first version (1.1), many ideas about the syntactic extensions of .bib files were already included, but the notions related to the bibliography styles and the implementation are now obsolete;
  - [5] summarises the experience we gained by implementing the first version in C, from a point of view related to re-engineering;
  - [6] shows the first relationship between MlBibTeX and XML's world, it has been reused when Version 1.3 was designed (cf. [9]);
  - [7] sketches the formal specifications that were used to help the first version's development—they were written using the CASL[13] language [1]—;
  - [12] explains why we started a second implementation of Version 1.3 in Scheme after an unfinished implementation in C.

- Last but not least, the following references describe some features planned but not implemented yet:
  - [8] describes a compatibility mode that allows bibliography styles written in the 'old' bst language to be interpreted by MlBibTeX[14];
  - [11] aims to put into action a framework for the definition of lexicographic order relations, such relations being language-dependent; presently, only calls to external functions defined in Scheme allow such definitions;
  - [13] shows a method, based on *patterns*, for removing LaTeX commands within values associated with BibTeX fields, they are transformed into strings or XML elements[15]: such a feature could be useful for users who wish to generate bibliographies for document written using the ConTeXt format[16] or DocBook[17].

Acknowledgements

I am indebted to Maurizio Himmelmann for all the fragments of this article written in the Italian language. In addition, he was waiting for this article for

[13] **C**ommon **A**lgebraic **S**pecification **L**anguage. This language has been designed by an international workgroup aiming to join efforts from different teams, in order to produce a unified algebraic specification language.

[14] This is the only part programmed in C in the first prototype of Version 1.3, and not re-implemented yet in Scheme (cf. [12]).

[15] See also [18]).

[16] To be fit for use, the definitions provided by TeX's core need to be organised in a *format*. The first format, developed by Donald E. Knuth, is *plain TeX* [21]; other formats are LaTeX [22], defined by Leslie Lamport, $\mathcal{AMS}$-TeX—the American Mathematical Society's format—defined by Michael D. Spivak [27], and ConTeXt, defined by Hans Hagen [2].

[17] DocBook is an XML-based system for writing structured documents [32].

a long period of time, thanks to him for his patience. I also thank Javier Pérez Campo, who helped me for the fragments written in Catalan and Spanish at Figure 1.

References

[1] Michel Bidoit, Peter D. Mosses, Till Mossakowski, Donald A. Sannella and Andrzej Tarlecki: casl *User Manual. Introduction to Using the Common Algebraic Specification Language*. No. 2900 in lncs. Springer. 2004.

[2] Hans Hagen: *ConTEXt, the Manual*. November 2001. `http://www.pragma-ade.com/general/manuals/cont-enp.pdf`.

[3] Jean-Michel Hufflen : « Vers une extension multilingue de BibTEX ». *Cahiers* gut*enberg*, Vol. 39–40, p. 23–38. In *Actes du Congrès GUTenberg 2001*, Metz. Mai 2001.

[4] Jean-Michel Hufflen: "MlBibTEX: a New Implementation of BibTEX". In: Simon Pepping, ed., *EuroTEX 2001*, pp. 74–94. Kerkrade, The Netherlands. September 2001.

[5] Jean-Michel Hufflen: "Lessons from a Bibliography Program's Reimplementation". In: Mark van den Brand and Ralf Lämmel, eds., ldta *2002*, Vol. 65.3 of entcs. Elsevier, Grenoble, France. April 2002.

[6] Jean-Michel Hufflen: "Multilingual Features for Bibliography Programs: From xml to MlBibTEX". In: Karl Berry and Kaja P. Christiansen, eds., *EuroTEX 2002*, pp. 46–59. Bachotek, Poland. April 2002.

[7] Jean-Michel Hufflen: "Using Algebraic Specifications to Formalise Multilingual Features". In: Martin Wirsing, Dirk Pattinson and Rolf Hennicker, eds., *WADT 2002*. Frauenchiemsee, Germany. Extended Abstract. September 2002.

[8] Jean-Michel Hufflen: "Mixing Two Bibliography Style Languages". In: Barrett R. Bryant and João Saraiva, eds., ldta *2003*, Vol. 82.3 of entcs. Elsevier, Warsaw, Poland. April 2003.

[9] Jean-Michel Hufflen: "European Bibliography Styles and MlBibTEX". tug*boat*, Vol. 24, no. 3, pp. 489–498. EuroTEX 2003, Brest, France. June 2003.

[10] Jean-Michel Hufflen: "MlBibTEX's Version 1.3". tug*boat*, Vol. 24, no. 2, pp. 249–262. July 2003.

[11] Jean-Michel Hufflen: "Order Relations and European Languages. Extended Abstract". In: Peter D. Mosses, ed., wadt *2004*, pp. 53–54. Barcelona, Spain. March 2004.

[12] Jean-Michel Hufflen: "A Tour around MlBibTEX and Its Implementation(s)". *Biuletyn* gust, Vol. 20, pp. 21–28. In *BachoTEX 2004 conference*. April 2004.

[13] Jean-Michel Hufflen: "MlBibTeX: beyond LaTeX". In: Apostolos Syropoulos, Karl Berry, Yannis Haralambous, Baden Hugues, Steven Peter and John Plaice, eds., *International Conference on TeX, xml, and Digital Typography*, Vol. 3130 of lncs, pp. 203–215. Springer, Xanthi, Greece. August 2004.

[14] Jean-Michel Hufflen: "Making MlBibTeX Fit for a Particular Language. Example of the Polish Language". *Biuletyn* gust, Vol. 21, pp. 14–26. 2004.

[15] Jean-Michel Hufflen: "Bibliography Styles Easier with MlBibTeX". In: *EuroTeX 2005 conference, program and preprints*, pp. 106–119. Pont-à Mousson, France. March 2005.

[16] Jean-Michel Hufflen: "MlBibTeX in Scheme (*First Part*)". *Biuletyn* gust, Vol. 22, pp. 17–22. In *BachoTeX 2005 conference*. April 2005.

[17] Jean-Michel Hufflen: *Managing Languages within MlBibTeX*. To appear. June 2005.

[18] Jean-Michel Hufflen: "Implementing a Bibliography Processor in Scheme". In: J. Michael Ashley and Michel Sperber, eds., *Proc. of the 6th Workshop on Scheme and Functional Programming*, Vol. 619, pp. 77–87. Indiana University Computer Science Department, Tallinn. September 2005.

[19] Richard Kelsey and William D. Clinger, eds., Jonathan A. Rees, Harold Abelson, Norman I. Adams iv, David H. Bartley, Gary Brooks, R. Kent Dybvig, Daniel P. Friedman, Robert Halstead, Chris Hanson, Christopher T. Haynes, Eugene Edmund Kohlbecker, Jr, Donald Oxley, Kent M. Pitman, Guillermo J. Rozas, Guy Lewis Steele, Jr, Gerald Jay Sussman and Mitchell Wand: "Revised^5 Report on the Algorithmic Language Scheme". hosc, Vol. 11, no. 1, pp. 7–105. August 1998.

[20] Oleg E. Kiselyov and Kirill Lisovsky: "xml, XPath, xslt Implementations as sxml, SXPath, and sxslt". In: *International Lisp Conference 2002*. San Francisco, California. October 2002.

[21] Donald Ervin Knuth: *Computers & Typesetting. Vol. A: the TeXbook*. Addison-Wesley Publishing Company, Reading, Massachusetts. 1984.

[22] Leslie Lamport: *LaTeX. A Document Preparation System. User's Guide and Reference Manual*. Addison-Wesley Publishing Company, Reading, Massachusetts. 1994.

[23] Frank Mittelbach and Michel Goossens, with Joannes Braams, David Carlisle, Chris A. Rowley, Christine Detig and Joachim Schrod: *The LaTeX Companion*. 2nd edition. Addison-Wesley Publishing Company, Reading, Massachusetts. August 2004.

[24] Oren Patashnik: *Designing BibTeX Styles*. February 1988. Part of BibTeX's distribution.

[25] Oren Patashnik: *BibTEXing*. February 1988. Part of BibTeX's distribution.

[26] Erik T. Ray: *Learning* xml. O'Reilly & Associates, Inc. January 2001.

[27] Michael D. Spivak: *The Joy of TEX*. American Mathematical Society, Providence, RI. 1986.

[28] The Unicode Consortium: *The Unicode Standard Version 4.0*. Addison-Wesley. August 2003.

[29] Doug Tidwell: xslt. O'Reilly & Associates, Inc. August 2001.

[30] W3C: xml *Path Language (XPath). Version 1.0*. w3c Recommendation. Edited by James Clark and Steve DeRose. November 1999. `http://www.w3.org/TR/1999/REC-xpath-19991116`.

[31] W3C: xsl *Transformations (*xslt*). Version 1.0*. w3c Recommendation. Edited by James Clark. November 1999. `http://www.w3.org/TR/1999/REC-xslt-19991116`.

[32] Norman Walsh and Leonard Muellner: *DocBook: The Definitive Guide*. O'Reilly & Associates, Inc. October 1999.