KDMiLe - Symposium on Knowledge Discovery, Mining and Learning

S. de Amo[1], A. P. L. de Carvalho[2], A. Plastino[3]

[1] Universidade Federal de Uberlândia, Brazil
deamo@ufu.br
[2] Universidade de São Paulo, Brazil
andre@icmc.usp.br
[3] Universidade Federal Fluminense, Brazil
plastino@ic.uff.br

Abstract. The Symposium on Knowledge Discovery, Mining and Learning (KDMiLe) aims at integrating researchers, practitioners, developers, students and users to present theirs research results, to discuss ideas, and to exchange techniques, tools, and practical experiences related to the Data Mining and Machine Learning areas.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms**.

Keywords: kdmile, template, data mining, machine learning

1. CALL FOR CONTRIBUTIONS

IMPORTANT: TITLE, ABSTRACT AND KEYWORDS MUST BE WRITTEN IN ENGLISH. Remaining sections of the paper may be written in english or portuguese.

KDMiLe is organized alternately in conjunction with The Brazilian Conference on Intelligent Systems (BRACIS) and the Symposium on Database Systems (SBBD). The KDMiLe Program Committee invites submissions containing new ideas and proposals, and also applications, in the Data Mining and Machine Learning areas. Submitted papers will be reviewed based on originality, relevance, technical soundness and clarity of presentation.

1.1 Scope and Topics

KDMiLe welcomes articles on a full range of research and applications on data mining and machine learning, including (but not limited to):

(1) **Data Mining Topics**
   —Association Rules
   —Classification
   —Clustering
   —Data Mining Aplications
   —Data Mining Foundations
   —Evaluation Methodology in Data Mining
   —Feature Selection and Dimensionality Reduction
   —Graph Mining

—Massive Data Mining
—Multimedia Data Mining
—Multirelational Mining
—Outlier Detection
—Parallel and Distributed Data Mining
—Pre and Post Processing
—Ranking and Preference Mining
—Privacy and Security in Data Mining
—Quality and Interest Metrics
—Recommender Systems based on Data Mining
—Sequential Patterns
—Social Network Mining
—Stream Data Mining
—Text Mining
—Time-Series Analysis
—Visual Data Mining
—Web Mining

(2) **Machine Learning Topics**
—Active Learning
—Bayesian Inference
—Case-Based Reasoning
—Cognitive Models of Learning
—Constructive Induction and Theory Revision
—Cost-Sensitive Learning
—Deep Learning
—Ensemble Methods
—Evaluation Methodology in Machine Learning
—Fuzzy Learning Systems
—Inductive Logic Programming and Relational Learning
—Kernel Methods
—Knowledge-Intensive Learning
—Learning Theory
—Machine Learning Applications
—Meta-Learning
—Multi-Agent and Co-Operative Learning
—Natural Language Processing
—Online Learning
—Probabilistic and Statistical Methods
—Ranking and Preference Learning
—Recommender Systems based on Machine Learning
—Reinforcement Learning
—Semi-Supervised Learning
—Supervised Learning
—Self-Supervised Learning
—Unsupervised Learning

1.2   Types of Submission

KDMiLe welcomes *research* articles that both lay theoretical foundations and provide new insights into the aforementioned areas and also *application* articles proposing to adapt existing data mining and machine learning technologies in some specific context, innovative commercial implementations

as well as reports and analysis of experiences in applying recent research advances in data mining and machine learning to practical situations.

1.3 Submission Instructions

Papers may be written in Portuguese or English, and must not exceed 8 pages. Papers exceeding this limit will be automatically rejected without being reviewed by the Program Committee. In addition, papers must be submitted in PDF format. Formats other than PDF will NOT be accepted. Papers exceeding this limit will be automatically rejected without being reviewed by the Program Committee. In addition, papers must be submitted in PDF format. Formats other than PDF will NOT be accepted.

1.4 Format Instructions

This folder has the following files:

—kdmileb.bib - an example of bibliography entries

—kdmile.cls - the latex template class for kdmile

—kdmile.bst - the latex template for bibliography entries

—template-kdmile.tex - an example of tex file

—template-kdmile.pdf - an example of the output pdf file

Prospective authors should take a look at the *template-kdmile.tex* for more information on the format, for example: how to add more than one institution to the affiliation and references to books [Baeza-Yates and Ribeiro-Neto 1999], book chapters [Borgida et al. 2009], conference papers [Ferreira et al. 2009], conference papers on book series (e.g., LNCS) [da Silva et al. 1996], journal articles [Laender et al. 2009], PhD thesis [Moro 2007] and information published online[1] [Berglund et al. 2007]. The file *kdmileb.bib* has examples of how to name conference proceedings and journals. In addition, Section 2 overviews the most important Latex instructions and is a "must-read" for authors with little or no experience in Latex, and Section 3 presents the most common errors.

2. LATEX INSTRUCTIONS

This section overviews some basic instructions for writing an article to KDMiLe using LATEX.

FIRST LINES

The first lines of your tex file identify the document class, packages, new definitions and environments the article will use. Specifically:

(1) The document class must be *kdmile.cls* and is defined as follows:

```
\documentclass[kdmile,a4paper]{kdmile}
```

Note that kdmile is published on A4 paper.

(2) You may use as many packages as you want. However, we suggest you use *graphicx* and *url* for figures and url format, and *fontenc* for avoiding warnings such as "*LaTeX Font Warning: Font shape 'OMS/cmtt/m/n' undefined*". You can define them as follows:

---

[1]Note: in case the URL refers to a software, dataset source or any other information that does not have a proper author, it is better to include it as a footnote, e.g.: "Biblioteca Digital Brasileira de Computação: http://www.lbd.dcc.ufmg.br/bdbcomp"

```
\usepackage{graphicx,url}
\usepackage[T1]{fontenc}
```

Please note that including the package *cite* may conflict with our template for bibliographic references (*kdmile.bst*).

(3) You may also use new definitions, such as the ones included in kdmile template:

```
\newtheorem{theorem}{Theorem}[section]
\newtheorem{conjecture}[theorem]{Conjecture}
\newtheorem{corollary}[theorem]{Corollary}
\newtheorem{proposition}[theorem]{Proposition}
\newtheorem{lemma}[theorem]{Lemma}
\newdef{definition}[theorem]{Definition}
\newdef{remark}[theorem]{Remark}
```

(4) You may also add new environments, such as the one included in kdmile template:

```
\newenvironment{latexcode}
{\vspace{0.1in}\setlength{\parindent}{18pt}}
{\vspace{0.1in}}
```

MANDATORY FIELDS

The initial lines are followed by mandatory fields which compose the article metadata. It is very important that all of them be correctly written as follows.

(1) Each article has a left and right-headers composed of two parts: the abbreviated names of the authors and the title. Both parts are defined as follows:

```
\markboth{A. Plastino and S. de Amo and A. P. L. de Carvalho}{KDMiLe - Symposium
on Knowledge Discovery, Mining and Learning}
```

Note that the first parameter has the authors, which are defined following these rules:
—If there is one author, then use author's full name, such as *Alexandre Plastino*.
—If there are two or three authors, then abbreviate each author's first name such as *A.Plastino and S. de Amo and A.P.L. de Carvalho*.
—If there are more than three authors, then the format is *A. Plastino et. al.*.
The second parameter is the title. If the title is too long, contract it by omitting subtitles and phrases, *not* by abbreviating words.

(2) The title of the article is defined as follows:

```
\title{KDMiLe - Symposium on Knowledge Discovery, \\ Mining and Learning}
```

Note that if you want to break the title in two or more lines, just add a line within it with two consecutive backlashes (\\).

(3) After the title, you need to define the name of the authors and their affiliations, as follows.

```
\author{A. Plastino, V. Braganholo}
\institute{Universidade Federal Fluminense, Brazil \\
\email{\{plastino,vanessa\}@ic.uff.br}
```

Note that both authors belong to the same university. If you have authors from different places, you write their information as follows:

```
\author{S. de Amo\inst{1}, A. P. L. de Carvalho\inst{2}}
```

```
\institute{Universidade Federal de Uberlândia, Brazil \\
\email{deamo@ufu.br}
\and
\institute{Universidade de São Paulo, Brazil \\
\email{andre@icmc.usp.br}
```

(4) The next step is to add the abstract within its environment, as follows.

```
\begin{abstract}
Text of your abstract using from 100 to 300 words, without any references.
Also, do avoid breaking it into paragraphs.
\end{abstract}
```

(5) After the abstract, you should include information for the ACM Computing Classification. The ACM Computing Classification System — `https://www.acm.org/publications/class-2012` — is a set of classifiers and concepts that describe the computing discipline. Authors can select entries from this classification system, via `https://dl.acm.org/ccs/ccs.cfm`, and generate the commands to be included in the LaTeX source.

(6) Then, the keywords are specified in alphabetical order as follows.

```
\keywords{kdmile,template}
```

Note that keywords can also be expressions, such as *Data Mining, Machine Learning*.

(7) The article then begins as follows.

```
\begin{document}
```

(8) It is common to thank funding agencies for research grants. If necessary, that information is defined in the *bottomstuff* environment, as follows.

```
\begin{bottomstuff}
This work was partially funded by CNPq grant number XYZ.
\end{bottomstuff}
```

(9) All the previous items complete the initial part of the article. Then, you need to use the *maketitle* command, which generates and formats all the previous items, as follows.

```
\maketitle
```

BODY CONTENT

The article then follows with its regular content with sections and subsections as follows.

```
\section{Section title}
```
which defines a section in first level (i.e., *1. Section title*).

```
\subsection{Subsection title}
```
which defines a section in second level (i.e., *1.1. Subsection title*).

`\subsubsection{Subsubsection title}` defines a section in third level (i.e., *1.1.1. Subsubsection title*).

For further information, we refer to the following web sources:

```
--http://en.wikibooks.org/wiki/LaTeX
```

```
--http://www.latex-project.org/
```

FINISHING YOUR ARTICLE

At the end of your article, you may add appendix sections as follows.

```
\appendix
\section{Title of first appendix section}
Text text text
```

After the appendix, you may also add a section for acknowledgments, as follows. Note that funding agencies should be mentioned as defined by the aforementioned environment $\backslash bottomstuff$.

```
\begin{ack}
Any acknowledgment you want.
\end{ack}
```

Then, you add the dates (if you do not add these lines, the footer of the title page will be wrong), as follows:

```
\begin{received} \end{received}
```

Finally, you must refer to the bibliography style and file, as follows.

```
\bibliographystyle{kdmile}
\bibliography{kdmileb}
```

where the file *kdmileb.bib* has all references. We try to keep the names of proceedings and journals following the same pattern. All references must be defined in the bib file, as follows.

```
@MISC{xpath,
    author = {Anders Berglund and Scott Boag and Don Chamberlin and
    Mary F. Fern\'{a}ndez and Michael Kay and Jonathan Robie and
    J\'{e}r\^{o}me Sim\'{e}on},
    title = {{XML Path Language (XPath) 2.0, W3C Recommendation}},
    howpublished = {http://www.w3.org/TR/xpath20},
    year = {2007} }
```

```
@INCOLLECTION{BorgidaCL09,
    chapter = {{Logical Database Design:  from Conceptual to Logical Schema}},
    pages = {1645-1649},
    title = {{Encyclopedia of Database Systems}},
    publisher = {Springer},
    year = {2009},
    editor = {Liu Ling and Tamer M. \"{O}zsu},
    author = {Alexander Borgida and Marco A. Casanova and Alberto H. F. Laender},
```

```
        address = {Berlin} }


@INPROCEEDINGS{FerreiraGALV09,
        author = {Anderson A. Ferreira and Marcos Andr{\'e} Gon\c{c}alves and Jussara
        M. Almeida and Alberto H. F. Laender and Adriano Veloso},
        title = {{SyGAR - A Synthetic Data Generator for Evaluating Name Disambiguation
        Methods}},
        booktitle = ecdl,
        year = {2009},
        pages = {437-441},
        address = {Corfu, Greece} }


@ARTICLE{LaenderMNM09,
        author = {Alberto H. F. Laender and Mirella M. Moro and Cristiano Nascimento
        and Patr\'{\i}cia Martins},
        title = {{An X-ray on Web-available XML Schemas}},
        journal = sigmodrecord,
        year = {2009},
        volume = {38},
        pages = {37-42},
        number = {1} }


@PHDTHESIS{Moro07,
        author = {Mirella M. Moro},
        title = {{The Role of Structural Aggregation for Query Processing over XML Data}},
        school = {University of California, Riverside (UCR)},
        year = {2007},
        address = {USA} }


@INCOLLECTION{SilvaLC96,
        chapter = {{An Approach to Maintaining Optimized Relational Representations
        of Entity-Relationship Schemas}},
        pages = {292-308},
        title = {Conceptual Modeling - ER'96},
        publisher = {Springer},
        year = {1996},
        editor = {Bernhard Thalheim},
        author = {Altigran Soares da Silva and Alberto H. F. Laender and Marco A. Casanova},
        volume = {1920},
        series = {Lecture Notes in Computer Science},
        booktitle = er }
```

3. MOST COMMON ERRORS

This section lists the most common errors when writing your paper. Note that *any* of these errors will certainly make the editors contact you for correction.

—Files do not compile. Editors will create the pdf files from the tex sources you submitted.
—Refering to work as "paper": you should refer to your work as article.

—kdmile metadata different from pdf file: author names must be exactly the same in the pdf as in the metadata at kdmile website [including abbreviations and uppercase letters].

—Generic CCS Concepts: in the item "CCS Concepts", the more specific the concept the better. Identify the lowest branches of the tree that seem to apply to your particular paper.

—Wrong header: with two or three authors, the header is as "A. Plastino and S. de Amo and A. P. L. de Carvalho"; otherwise, it is "A. Plastino et. al." (use $\markboth$).

—Acknowledgments as section/subsection: acknowledgments must be in the appropriate place (using environment $\bottomstuff$).

—Long/wrong abstract: abstract should have only one paragraph; in the kdmile metadata, it cannot have tex commands.

—Incomplete/wrong references: *all* references must strictly follow kdmile template, i.e., conference/workshop papers with complete authors, title, proceedings name, year, paper initial and final pages, and city/country address (no series, no url, no date, no total number of pages, no publisher - see template example); journal articles must have authors, title, journal name, pages, both volume and number; double check the template for conference papers published on Lecture Notes; webpage can be footnote instead of reference.

—Undefined references.

—Style: try to avoid orphan words (one word alone in one line), orphan lines (one line alone at the beginning of a page), single subsections (e.g., section 3, subsection 3.1, section 4); place figures and tables at the top [t] or bottom [b] of your pages.

—File compiles with warnings: try to eliminate all warnings, they may affect the final pdf.

—Wrong characters: words in Portuguese (even inside bib file) must have proper characters (e.g., Computa\c{c}\~{a}o).

—Outdated template: check the website for the updated template (which corrected some errors from the previous one).

REFERENCES

Baeza-Yates, R. A. and Ribeiro-Neto, B. A. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.

Berglund, A., Boag, S., Chamberlin, D., Fernández, M. F., Kay, M., Robie, J., and Siméon, J. XML Path Language (XPath) 2.0, W3C Recommendation. http://www.w3.org/TR/xpath20, 2007.

Borgida, A., Casanova, M. A., and Laender, A. H. F. Encyclopedia of Database Systems. Springer, Berlin, Logical Database Design: from Conceptual to Logical Schema, pp. 1645–1649, 2009.

da Silva, A. S., Laender, A. H. F., and Casanova, M. A. Conceptual modeling - er'96. In *Proceedings of the International Conference on Conceptual Modeling*, B. Thalheim (Ed.). Lecture Notes in Computer Science, vol. 1920. Springer, An Approach to Maintaining Optimized Relational Representations of Entity-Relationship Schemas, pp. 292–308, 1996.

Ferreira, A. A., Gonçalves, M. A., Almeida, J. M., Laender, A. H. F., and Veloso, A. SyGAR - A Synthetic Data Generator for Evaluating Name Disambiguation Methods. In *Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries*. Corfu, Greece, pp. 437–441, 2009.

Laender, A. H. F., Moro, M. M., Nascimento, C., and Martins, P. An X-ray on Web-available XML Schemas. *SIGMOD Record* 38 (1): 37–42, 2009.

Moro, M. M. *The Role of Structural Aggregation for Query Processing over XML Data*. Ph.D. thesis, University of California, Riverside (UCR), USA, 2007.