Scientific documents written by novice researchers: A personal experience in Latin America

Ludger O. Suarez–Burgoa

Abstract

This article presents 20 years of the author's personal experience — described as a particular Latin American experience — about the elaboration of scientific documents created by novice researchers, from the use of the typewriter to prepare a school scientific report up to the conception of a TeX-family class file (i.e. the `unbDscThesisEng` class of the *Universidade de Brasília* Doctoral Thesis, English Version) to prepare his theses. He also gives opinions about a possible Universal Editable Format for scientific documents made by novice researchers, to allow such documents to persist over time without losing information due to changing encoding formats of proprietary software.

1 Introduction

Processing scientific documentation is an essential part of publishing results. It is a concern in scientific institutions and academia, because it requires text, tables, equations, figures, and references in a relational structured and compact manner; and all these in conjunction reflect the quality of the message desired to be presented.

Nowadays, researchers individually or in small groups (called in this article *novice researchers*, NRs) have also become actors in this important process, because they can disseminate their results in various effective media now available: indexed journals, both printed and electronic, conference articles, Internet links, and blogs, for example.

With the general availability of free and open source software (FOS) and with increasingly abundant information on the Internet with details and recommendations for proper use of these FOS tools, NRs can now present fully-developed, high-quality, and well-formatted scientific documents. They can be part of the development of scientific documentation by creating support for particular institutions (e.g. TeX-family class files for university theses).

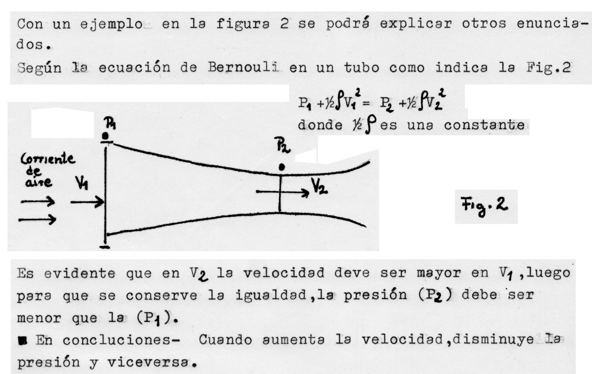
In the following sections the particular experience of an NR of Latin America will be discussed, who has prepared scientific documents and technical reports for around 20 years in Bolivia, Colombia, Brazil, Argentina, Perú, and Chile; and now feels comfortable using: TeX for text, tables, and equation creation, editing and management; SVG to create and edit graphics and plots; and BibTeX to store and manage references of scientific documents.

2 Some particular past experiences

Most of the Latin American NRs who nowadays are writing doctoral theses have suffered in the transition from the typewriter to computer software, when dealing with scientific and technical documents.

For example, for a technical note at a school science fair, it was common to see in the 1980s the text and some lines of the document prepared with a mechanical typewriter (e.g. an Olympia AG coming from Wilhelmshaven, Germany, with a two-color ink tape: black and red), and graphics, sketches and formulas executed by hand. In special cases, one might have the privilege of using an electronic typewriter (e.g. a Brother CE-30, made in Taiwan), which offered the possibility of correcting some mistakes before typing on the paper. In this stage, one used photocopies for mass reproduction; therefore, graphics should be conceived only in black and white (B&W) and should be drawn on good paper with ink.

The artistic part of this was to use different widths and types for lines, different textures for fills, and different font sizes. Figure 1 shows a part of the author's first scientific document, made in 1992 [6]. Observe that the black color of the letters is not uniform, being dependent on the force with which one's finger triggered the letter key and the ink tape quality or usage. Also observe the equation with superscripts, subscripts, and Greek letters done by hand; and the graphic was also done by hand.

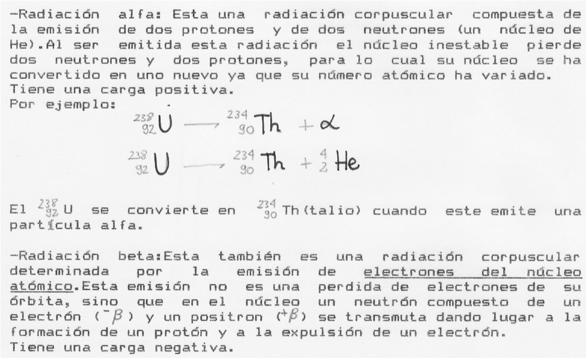

Con un ejemplo en la figura 2 se podrá explicar otros enunciados.
Según la ecuación de Bernouli en un tubo como indica la Fig.2

$P_1 + ½\rho V_1^2 = P_2 + ½\rho V_2^2$
donde ½ρ es una constante

Es evidente que en V_2 la velocidad deve ser mayor en V_1, luego para que se conserve la igualdad, la presión (P_2) debe ser menor que la (P_1).
■ En concluciones- Cuando aumenta la velocidad, disminuye la presión y viceversa.

Figure 1: Part of a document presented at a science fair, done with typewriter and by hand [6]

A few novice researchers had first contact with a computer word processor in the late 1980s, e.g. WordStar 4.0 under the CP/M operating system. But, in the early 1990s — as the present author was finishing secondary school age and in the initial stages of a university bachelor's course — some NRs had the opportunity to use *friendly* word processors on personal computers (e.g. Word Perfect 5.1 under the Disk Operating System (DOS) version 4.0).



-Radiación alfa: Esta una radiación corpuscular compuesta de la emisión de dos protones y de dos neutrones (un núcleo de He).Al ser emitida esta radiación el núcleo inestable pierde dos neutrones y dos protones, para lo cual su núcleo se ha convertido en uno nuevo ya que su número atómico ha variado.
Tiene una carga positiva.
Por ejemplo:

$$^{238}_{92}U \longrightarrow ^{234}_{90}Th + \alpha$$

$$^{238}_{92}U \longrightarrow ^{234}_{90}Th + ^{4}_{2}He$$

El $^{238}_{92}U$ se convierte en $^{234}_{90}Th$ (talio) cuando este emite una partícula alfa.

-Radiación beta:Esta también es una radiación corpuscular determinada por la emisión de electrones del núcleo atómico.Esta emisión no es una perdida de electrones de su órbita, sino que en el núcleo un neutrón compuesto de un electrón ($^{-}\beta$) y un positron ($^{+}\beta$) se transmuta dando lugar a la formación de un protón y a la expulsión de un electrón.
Tiene una carga negativa.

Figure 2: Part of the document presented to another science fair, done with Word Perfect 5.1 under DOS 4.0 and by hand [7]

The problems of this system (both hardware and software) were that in the region only English-language keyboards were available, and the text processor was initially available only for English. Therefore, in order to use the system in Spanish language, the ñ and vowels with accent (e.g. á, ó, í) had to be introduced by a combination of the Alt key plus the appropriate three numbers taken from the American Standard Code for Information Interchange (ASCII). So, it was common to see a laminated cheat sheet above the function keys on a keyboard.

Formulas and sketches were normally a combination usage of the word processor application tools — which were not good for complex sketches and formulas — and handwriting; and graphics could be prepared with other software (e.g. Harvard Graphics, which was not the proper tool, but it did manipulate vector graphics) or by hand. At this point, one began to use dot matrix printers, but they were very slow. For example, in order to print a 32-page document on a Wang Labs PM 016/160 dot matrix printer, one consumed five hours listening to the particular sound of those printers and taking care that no paper problems arose. Later, this printer type became faster in the region, e.g. the EPSON LQ-300.

Figure 2 shows an example from a document created in 1992 in the environment described above. Observe that even though the document was made with a word processor, the quality remains close to the document prepared two years ago (Figure 1) because it was hardware-dependent (i.e. on the printer type, which was a dot matrix printer that also uses an ink tape).

Also, observe in Figure 2 that some special characters in the Spanish language — for example the `í` letter — are printed with a different quality (e.g. the last but one word *partícula* of the paragraph after

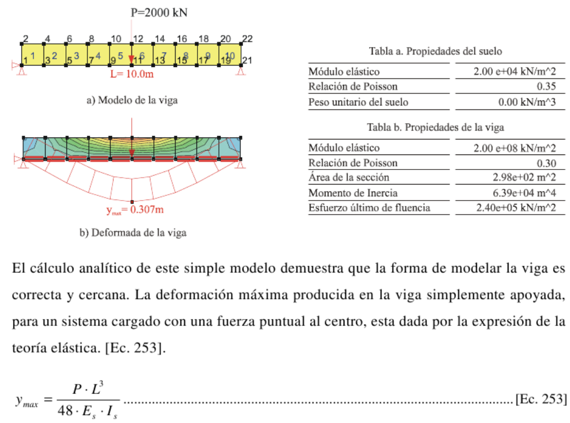


Tabla a. Propiedades del suelo

Módulo elástico	2.00 e+04 kN/m^2
Relación de Poisson	0.35
Peso unitario del suelo	0.00 kN/m^3

Tabla b. Propiedades de la viga

Módulo elástico	2.00 e+08 kN/m^2
Relación de Poisson	0.30
Área de la sección	2.98e+02 m^2
Momento de Inercia	6.39e+04 m^4
Esfuerzo último de fluencia	2.40e+05 kN/m^2

El cálculo analítico de este simple modelo demuestra que la forma de modelar la viga es correcta y cercana. La deformación máxima producida en la viga simplemente apoyada, para un sistema cargado con una fuerza puntual al centro, esta dada por la expresión de la teoría elástica. [Ec. 253].

$$y_{max} = \frac{P \cdot L^3}{48 \cdot E_s \cdot I_s} \quad \text{[Ec. 253]}$$

Figure 3: Part of the BSc thesis, this done with MS Word under Windows 3.1 and Graphics with CorelDraw 9 exported to a medium or low raster JPG format [8]

the equations). Also, the equations

$$^{238}_{92}\text{U} \rightarrow^{234}_{90} \text{Th} + \alpha \tag{1a}$$

$$^{238}_{92}\text{U} \rightarrow^{234}_{90} \text{Th} +^{4}_{2}\text{He} \tag{1b}$$

and in-text variable symbols (e.g. `...un positrón` ($^{+}\beta$) `se transmuta...`) still needed human intervention.

At the final stages of the university bachelor course, at the end of the 1990s, things went better: the Windows 3.1 (or higher) Operating System (OS) in Spanish came with the not-well-known — at that time — MS Word text processor, and now keyboards were available for Spanish. Also, this text processor included programs to prepare sketches and formulas, which were more flexible, allowed more complex cases, and, best of all: they supported full color. Also, because color bubble jet printers were accessible to an NR, one started to conceive of full-color figures in scientific and technical documents.

Figure 3 shows the full-color graphics (grayscaled for *TUGboat* hardcopy, though) made in a vector program, which decreased in quality when exporting to a raster format; but this was necessary because the text processor did not import properly and exactly the given graphic, even when trying to use its own EPS vector format. The equation editor of the word processor allowed these simple equations to be executed well.

For the first decade of this new century, use of these tools has been accepted by some scientific researchers and most industrial professionals, with the difference being that there was more diversity of fonts, document templates, and printing resources. This happened more due to hardware improvements (rather than improvements in the commercial software), which now allowed storing huge amounts of data in memory during editing; therefore, one could insert in a What You See Is What You Get (WYSIWYG) document file — for example — high resolution raster figures and be unlikely to suffer a program crash and consequently suffer a file corruption.

Nowadays, the quality of hardcopy documents has been improved, because laser color printers are more accessible for NRs than in the past; therefore, more color texts have been produced and they can also be reproduced economically in small quantities.

For example, most Master's degree dissertations in civil engineering in Latin American universities have been prepared with the combination of proprietary WYSIWYG software (i.e. word processor, spreadsheet, and vector graphics editor), known as *office* programs. Figure 4 shows part of an MSc thesis prepared with these programs: reference citations, reference lists, and lists of variables, abbreviations, and acronyms were handled manually; on the other hand, figure and table referencing, and equation numbering, were created automatically.

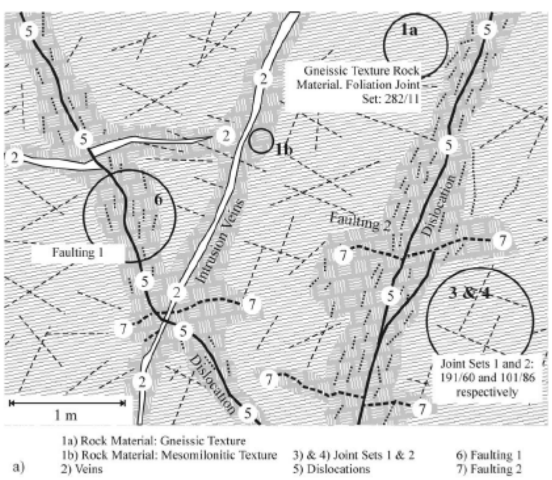


FIGURE 93. Rock Mass Describing Their Main Six Structures
a) In a wall.

Massivity Parameter (f_σ) gives the grade of massiveness and continuity of the rock mass. In order to characterize a rock mass as massive and continuous, the parameter f_σ should be greater or equal to 0,5 (Palmström, 2003).

$$J_C = \left(\frac{J_R}{J_A}\right) \cdot J_L \qquad (70.)$$

For P3UHC rock mass, the massivity parameter was encountered equal to 0,44 (Table 58). That means, under RMi concept, that the ratio between the uniaxial compressive strength of rock mass and the uniaxial compressive strength of the rock material for P3UHC is equal to

Figure 4: Part of an MSc thesis, done with MS Office under Windows XP and graphics with CorelDraw X4 exported in high-resolution raster (JPG) format [9]

But the biggest drawback of using WYSIWYG programs was that users did not learn about document structuring. They used a text processor as a simple electronic typewriter with improved graphical capabilities. This situation is becoming less common in recent years; now, users are considering the importance of hierarchically structuring documents, with the goals of facilitating exporting to XML and of disseminating documents via the Internet.

2.1 Digital Dark Age also touches NR

Past documentation stored in an encrypted coding and unable nowadays to be reopened, because there is no access to the hardware and software with which it was created, has suffered a permanent loss — even when the stored data is well preserved. If this situation occurs in a period between the time when the first electronic document was stored without preventing this situ, to the time when the last electronic document was stored before this situation is solved, then that period so determined can be considered a Digital Dark Age (DDA), because nobody in the future will be able to know what humans had documented in that period.

Because humans have not in fact resolved this situation, we are now inside a DDA. On a small scale and for particular usage, NR are also suffering the consequences of DDA. For example, some plots of the scientific document shown in Figure 2 [7] were made using proprietary software (Harvard Graphics, as mentioned above), which stored the files with the extension CHT. To date, no emulator can visualize these graphics, even though the file is stored perfectly safely on a hard disk.

NRs being in a DDA also causes the loss of many of their old documents, which are frequently more vulnerable than most others. This is because most documents developed by NRs are not of global interest; therefore, those documents are not stored in data centers but instead on a home system. In 20 years of managing personal scientific documentation, the present author — an NR — has suffered two important losses of documentation: first in the year 2000, when the entire computer was stolen; the other in 2007, when the author's first old PC computer and the programs' floppy disks — preserved in order to minimize personal DDA — were erroneously sold by the author's father.

Nowadays the situation is becoming less problematic for NR, since large external companies offer well-implemented digital data storage alternatives (e.g. Dropbox).

2.2 The experience with the TEX family

TEX is a computer typography program, free and open source in today's terminology, created in 1978 by Donald E. Knuth of Stanford University, which has revolutionized digital typesetting for scientific publishing and transformed the process of putting mathematical ideas on paper [2, 3].

Since then, many improvements and proposals have followed from it, creating the so-called TEX family (also denoted (LA)TEX). In general, this family has shown in its time that it was designed with two main goals in mind: to allow anybody to produce high quality documents; and to provide a system that would give exactly the same results on all computers, now and in the future [4].

One advantage of (LA)TEX, among many others, is that one can use escape sequences for characters beyond basic ASCII; therefore, one can cover all possible special characters with only 128 characters defined in the OS. Even though this limit was superseded many years ago with improved OSs, the use of only ASCII characters can still be very useful when one wants to put document information in structured databases: for a word-searching process, it is better to use the least possible number of different characters, as a broader range can easily introduce more errors.

The personal experience of the author with TEX began in 1996 when a friend seriously illuminated the benefits of the new "free" OS Linux — erroneously conceived by the author as *free of charge* — because the text processor software it offered was LATEX-based. In those years, another friend that came from Switzerland to visit his native country (i.e. Bolivia) gave the author three or five $3\frac{1}{2}$ inch floppy disks with a program that dealt well with scientific documents (i.e. Scientific Workplace under Windows 3.0 OS). The present author tried to make use of this, but because tutorials were lacking, finally he abandoned the program.

Unfortunately, the author ignored this clever advice, and 14 years passed before he recognized that the pair GNU/Linux and TEX had an important ideology behind them (see for example [5]) and improved advantages for NRs.

3 Universal Editable Format

Unfortunately, novice researchers coming from Latin American countries generally still believe that encoded binary files coming from popular office packages are the correct candidate as a universal format for document editing. In the lexicon of individuals, one finds phrases such as "send me a DOC file, please" when one wants to receive a document.

Also, in many university libraries of Latin America, the format they have adopted as universal is, erroneously, MS Word DOC format in its 2003 version. Strangely, they do not at least adopt the MS Word DOCX format, which permits exporting DOCX to XML. A disadvantage of adopting DOCX is that one must run the proper program to do the transformation to XML, and therefore one also is still dependent on commercial software.

Other universities adopted Rich Text Format (RTF), but this does not support graphics, formulas, tables, etc; therefore, this solution is even worse than the first, and worse also than PDF.

In general terms, a Universal Editable Format (UEF) need not necessarily be known by everybody. Instead, it should meet the following basic requisites:

- persist across time without losing information (i.e. be 100% identical on all machines and avoid DDA).
- be freely accessible anytime and anywhere (i.e. not be proprietary);
- be independent from hardware;
- not be encrypted or represented using binary codes; therefore, be readable by any text editor, even the simplest one;
- have commands understandable by any individual after reading available free[1] documentation;
- the binary programs that parse the commands should run under any OS.

Particularly for an NR, the TEX family can be a serious candidate because:

- it has persisted for many years without important weaknesses;
- it is the oldest system that has been a reliable, free, de facto standard for decades;
- it has a great number of users; and
- it can be used by any individual, including those without enough money to buy commercial editing program licenses, but with access to a computer and the Internet (e.g. pre-university, bachelor, and postgraduate students, i.e. NRs).

3.1 Document interchange formats

Electronic documentation and dissemination was not commonly used by NRs in Latin America until the middle of the 1990s. Before the Portable Document File (PDF) format was devised, other formats were available in the world — for example the DeVice Independent format (DVI) developed as part of the TEX family, and the PostScript format coming from Adobe — these were not known by the novice researchers in this region.

It was only after the PDF format and the Acrobat reader came into prominence that people encountered a viable use for and storage of electronic

[1] Free in the sense of freedom

documentation. The interchange of documents nowadays in this region is handled well via this format. And, since 2008 this format has become even more popular, after it was *liberated* by Adobe, made an ISO standard and in general available for any individual to make, use, sell and distribute PDF-compliant implementations.[2] Since then, FOS software has been developed with considerable added value, for example, PDF editing and electronic signatures.

The PDF format can be a good option to preserve a document as it was conceived by its author(s) and for electronic libraries (i.e. static preservation); but perhaps is not the proper one to be adjustable to the flexibility of web pages, and for the more volatile and interactive areas of the web which require dynamic document preservation and/or continuing editing, as for example wiki pages (e.g. Wikipedia) [2]. Also, browsers cannot display PostScript or PDF files without the aid of extra software (i.e. add-ons, many of which are proprietary), which in turn causes many readers to choose to download such files and view them offline or print them.

In the near future — not to say *in the present* — portable devices (e.g. tablets, cell phones, smart phones) will be more popular, and programs *in the cloud* will be used more; therefore, it is possibly necessary to adopt another format for document interchange for this new technological tendency.

With various scripts or programs, documents created by TeX can be transformed to XML which can be a first choice candidate for an NR to have a good interchange document format meeting such Internet-related requirements.

3.2 The case of graphics

In the 1990s, one typically used so-called vector graphic editor programs (e.g. CorelDraw and Adobe Illustrator for Windows, which were and are commercial programs) to execute figures for documents. Unfortunately, in the early years of the Internet, information about the existence of FOS programs was lacking. Therefore, piracy of the most famous graphic programs was the common "solution" for NRs in Latin America during those years.

Other NRs erroneously used other sorts of proprietary programs to make graphics for their documents, for example presentation programs (e.g. PowerPoint) and two-dimensional computer aided design (CAD) programs (e.g. AutoCAD). This was typically because they had no choice other than to use the software they had available, legal or otherwise.

[2] But this donation by Adobe did not follow the free (as in freedom) or open software concepts, because the source code was not liberated.

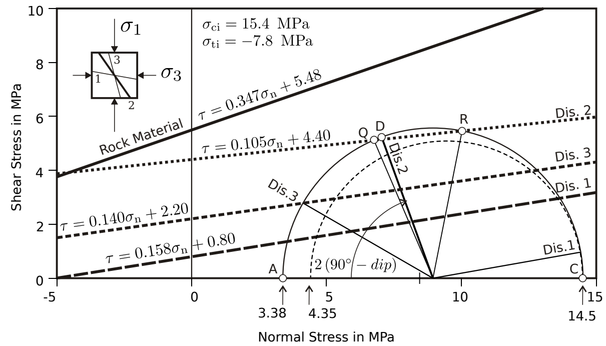


Figure 5: A graphic from the PhD thesis, created with Inkscape 0.47 under i486-PC-Linux-GNU (Debian Squeeze 4.4.5-8) and exported to PDF format [10]

One principal limitation found by NRs working this way was that they became dependent on proprietary file formats, which were — and still are — difficult to convert to a hypothetical universal editable format for 2D and 3D vector graphics. This might be the reason why some CAD proprietary file formats became so *important* for interchange of graphics (e.g. the AutoCAD DWG format); at the same time, they are a poor choice for interchange, because the structure of the DWG format regularly changes with new releases, and has never been made publicly available.

This situation continues today, but with fewer devotees. Nowadays, NRs more often use less complicated and costly proprietary software, or have decided on well-developed FOS alternatives.

With the advent of the new century, a new vector graphics specification was developed for public use: Scalable Vector Graphics (SVG), based on XML, which nowadays may be considered a universal vector editable format, being widely supported on the Internet and in applications. Unfortunately, SVG still supports only 2D graphics, but efforts are being made to cover 3D as well in the future.

SVG figures can be created and edited with any text editor, but it is often more convenient to create and edit them with a natural human interface for graphics. In the area of FOS software, one interesting program that deals with this format is the excellent Inkscape, which is available for all major OSs. Figure 5 shows a scientific figure developed entirely with this program using the SVG format; observe that TeX equations are included.

3.3 (LA)TeX classes as document managers

Returning to discussion of the TeX family, for many years the *class* concept has been used for (LA)TeX-based documentation, by creating a *class file* (CLS extension by default). Although most commonly known for LaTeX, the concept can be applied anywhere in the TeX family.

This file — if properly designed — permits an NR to create a document structure and formatting rules once (i.e. a document template), and then re-use this many times. This makes it possible to create uniform documents with respect to format and structuring, broadly used for book collections, journal articles, institutional reports, university dissertations and theses, and by extension any serialized document.

A CLS file is used under the TeX family concepts, formats and engines. Thus, its use can result in high quality document creation and a suitable document manager, since in a class file one can, for example:

- restrict the document formatting and the document structure;
- define the text and mathematics font styles;
- define table formatting;
- limit the number of chapters and appendixes;
- use a specific format for bibliographic references;
- define proper format for the frontmatter, main text, and backmatter; and
- define the page size, margins, headings and footers, and numbering.

The elaboration of a CLS file for (LA)TeX is not terribly difficult, but nor is it an easy task; by personal experience, it can be done by people who have used (LA)TeX for at least one year, with the help of the plethora of information on the Internet (a situation that was not possible 15 years ago). On the other hand, the use of an existing CLS file within (LA)TeX is a very easy procedure, which any NR beginner can do.

Therefore, as this is an excellent tool, the creation of any CLS file for any commonly used document at public institutions can be useful for others. For this reason, over the last few years (perhaps five years), the availability of CLS files for universities' documentation (i.e. dissertations and theses) has been increasing.

An observation about this last statement: such dissemination is not usually driven by the libraries or other university offices; instead, it is usually comes from the student community, with or without the aid of a professor (i.e. an NR). Regarding this, it is the mathematical departments of the universities that have typically promoted the use of (LA)TeX and who have created and disseminated CLS files.

Following this last-mentioned tendency, in the following section it is briefly presented how a CLS script file can be structured and the minimum environments it might provide. This CLS file was developed by the author for his doctoral thesis, and his general conclusion — among others mentioned in the next section — that emerged when working with the TeX family was that finally, after 20 years, he has found the proper tool for scientific documentation.

4 The unbDscThesisEng TeX family class

The `unbDscThesisEng` TeX family class is an unofficial template (i.e. not approved by the university) for the English language version thesis of the Geotechnical Postgraduate Program of the *Universidade de Brasília*, Brazil.[3] It is composed of a main CLS script file which requires other secondary files and a particular directory structure.

This class was based on the `book` class with modifications to accomplish local formatting requirements. It could also be properly used for the Portuguese language by doing some modifications not yet included in this version. The CLS file is free software that is released under the terms of GNU General Public License Version 3, as published by the Free Software Foundation.

The class permits the student to use the following eleven environments:

- `princover` for making the main cover of the thesis by attaching a `background.pdf` file;
- `maketitle`, a redefinition of the original in the `book` class, that permits making the title page according to the university rules and format;
- `approbationpage` for the page with the jury members' names and signatures;
- `catalogingpage` for the page of the thesis cataloging and copyright information;
- `dedicatory` for the dedicatory text;
- `acknowledgements` and `agradecimentos` for the acknowledgments in English and Portuguese;
- `abstract` and `resumo` for the abstract in English and Portuguese;
- `tableofcontents` for the TOC;
- `listoffigures` and `listoftables` for the lists of figures and tables;
- `listofabbrevsymbs` for the list of abbreviations and symbols (separated by Latin and Greek characters);
- `listofreferences` for the references after the main matter;
- `invitationpage` for a page with the date and time of the thesis presentation;
- `reportpage` for a page with general report information of the document.

In order to use any of the environments, the `unbDscThesisEng.cls` file should reside in the document's main directory, and 48 variables are available

[3] This class is at Version v1.0 as of 2012/15/08

for the user to fill. These variables are filled by the user in a separate TeX file (i.e. the `initials.tex` file), in order to avoid editing the CLS file.

Also, a main TeX file was designed (`aaaThesis.tex`), in which the user can: define the page size and its margins; define the text size among 10 pt to 12 pt; define if the document should be printed as two-side or single-sided; introduce other (LA)TeX packages; redefine some commands; and insert the document text. All the figures should be placed in a directory called `FIGURES`; and front-, main-, and backmatters should be inside the directories `FRONT_MTR`, `MAIN_MTR`, and `BACK_MTR`, respectively. The bibliographic references should be written in BibTeX format and be named `bibliography.bib`. The bibliographic style used for this class is `chicnarm.bst`, recommended to reside in the same document main directory.

Because this class uses the `nomenclature` package to manage the list of variables, a `nomencl.cfg` file should also be in the document main directory.

The following listing shows the directory structure for the `unbDscThesisEng` class and the minimal required files. The user interface used in this example listing is the KDE Integrated LaTeX Environment (KILE) under KDE Platform Version 4.4.5 for GNU/Linux; the `unbDscThesisEng.kilepr` file also defines the complete document structure.

```
drwx  BACK_MTR
drwx  FIGURES
drwx  FRONT_MTR
drwx  MAIN_MTR
-rw-  aaaThesis.nlo
-rw-  aaaThesis.nls
-rw-  aaaThesis.pdf
-rw-  aaaThesis.tex
-rw-  background.pdf
-rw-  bibliography.bib
-rw-  chicnarm.bst
-rw-  initials.tex
-rw-  invitationBackground.png
-rw-  nomencl.cfg
-rw-  nomencl.dtx
-rw-  nomencl.ins
-rw-  nomencl.ist
-rw-  unbDscThesisEng.cls
-rw-  unbDscThesisEng.kilepr
```

Experiences using this class for the doctoral thesis brought about improvements in: text formatting; correct use of SI units according to the proper rules; good mathematical symbolization of variables; robust index generation — especially with the variables index; robust reference structuring and full hypertextual citations. All of these in general improved the document in quality and also reduced the time needed for its elaboration.

Experiences using the SVG format for graphics in the thesis brought about: homogeneous text formatting in all figures; retaining in the graphics the same mathematical variables defined in the documents (by using the LaTeX equation rendering extension of Inkscape); easy editing independent from the text of the document; and high quality output, being vector graphics and not bitmaps.

5 Final comments

Comments given here are based on the author's personal experience in his region and language, having had the opportunity to deal with the scientific and technical document creation for some 20 years, and knowing the popular text processors and typesetting programs.

In the historical narration of the author, it appears that milestones of that theme (i.e. use of computer typesetting, use of DOS, Windows OS, bubble and laser jet printers, and other issues mentioned in Section 2) in Latin America came later than in other regions; but this is not accurate. The true reason is that the abovementioned milestones came relatively late to NRs, who are ordinary people: university students and teachers. It is likely that these milestones came sooner to the region in more specialized areas, for example, high-level research institutions.

The TeX family tries to reduce Digital Dark Age (DDA) hazards for the NR. Preventing the DDA for FOS formats could be less costly and time-consuming than doing the same for proprietary formats. The latter requires signing agreements with the format owners, who try to market their work by praising it as *social labor* when preventing DDA, while in fact they are part of the DDA problem; proprietary formats are by nature against DDA prevention.

What seems to be true is that the correct way forward in scientific document preparation for NRs is through the FOS concepts and by typesetting programs such as the TeX family. When mentioning this duality, FOS concepts & TeX family, it is inevitable to hear about the robust ideology they have behind them; and probably from this duality can emerge the proper UEF sought by the NR.

Perhaps the tendency of individuals in using SVG format for graphics, and TeX for text, tables, equations and references could mark a future tendency in using XML format for scientific documents (e.g. DocBook). But TeX is evidently easier to learn than XML, and at any rate, there exist proper ways to translate TeX to XML.

Also, in the near future, it is possible that the ePub format — a structured compressed XML based format broadly used for mobile devices — can be another possible interchange document format for the NR, rather than PDF. This can work because the Internet is a channel for distributing publications and preprints in many disciplines, as well as becoming a venue for less formal jottings and conversations using mobile technology.

At this moment, TeX has become the *de facto* standard text processing system in many academic high-level scientific and research institutions, while — in parallel — it is increasingly the choice of NRs in developing countries (e.g. Latin America). The example of usage of the `unbDscThesisEng` class is one of many found in literature which improved a scientific document conception, elaboration, and competitive dissemination, which has a good opportunity to prevail for many years without being caught by DDA. But TeX family classes should be better promoted by universities, who should also define more specific and rigorous document elaboration policies.

Finally, digital storage is easy but digital preservation is not. Preservation means keeping the stored information cataloged, accessible, and usable on current systems, which requires constant effort and expense. One cannot reverse the digitization of everything; what one has to do is convert the design of software from brittle to resilient, from heedlessly headlong to responsible, and from time-corrupted to time-embracing [1].

References

[1] S. Brand. Escaping the digital dark age. *Library Journal*, 124(2):46–49, February 1999.

[2] H. Brian. Writing math on the web. *American Scientist*, 97(2):98–102, March–April 2009.

[3] S. Ditlea. Rewriting the Bible in 0's and 1's. *Technology Review*, 102(5):66–70, September–October 1999.

[4] A. Gaudeul. Do open source developers respond to competition?: The (LA)TeX case study. *Social Science Research Network*, (908946), March 2006.

[5] L.E. Rosen. *Open Source Licensing: Software Freedom and Intellectual Property Law*. Prentice Hall, Upper Saddle River, NJ, 1st edition, 2004.

[6] L.O. Suarez-Burgoa. Estudio aerodinámico, fuerza de sustentación y resistencia en un avión. Technical report, Colegio Alemán Mariscal Braun, La Paz, Bolivia, October 1990.

[7] L.O. Suarez-Burgoa. Estudio nuclear de átomos radiactivos y efectos de la radiactividad. Technical report, Colegio Alemán Mariscal Braun, La Paz, Bolivia, October 1992.

[8] L.O. Suarez-Burgoa. Consideraciones para estabilizar taludes por medio de un sistema de tablestacado y enrejado vegetado. BSc. thesis, Universidad Mayor de San Andrés, La Paz, Bolivia, June 2001.

[9] L.O. Suarez-Burgoa. Rock mass mechanical behavior assessment at the Porce III underground hydropower central, Colombia, South America. MSc. thesis, Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia, February 2008.

[10] L.O. Suarez-Burgoa. *A qualitative physical modeling approach of rock mass strength*. Ph.D. thesis, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília DF, August 2012. Publication G.TD-079/12.

◇ Ludger O. Suarez–Burgoa
Universidad Nacional de Colombia
Fac. Min., Esc. Ing. Civ.
Cl. 65 #78 – 28, Bl. M1 Of. 320
Medellín, Colombia
`losuarezb (at) unal dot edu dot co`
`http://geomecanica.org/`