POLITÉCNICO DE LEIRIA

Improving Machine Learning Efficiency Against Noisy Data Sources

Investigating Advanced Strategies to Mitigate Adverse Effects of Noisy Data

Joe Smith

School of Management and Technology

Department of Computer Engineering

Master in Cybersecurity & Digital Forensics

Leiria, August 2024

This page intentionally left blank.

POLITÉCNICO DE LEIRIA

Improving Machine Learning Efficiency Against Noisy Data Sources

Investigating Advanced Strategies to Mitigate Adverse Effects of Noisy Data

Joe Smith

Supervisor: John Smith
Full Professor, Polytechnic of Leiria

Co-supervisor: Steve Smith
Associate Professor, Polytechnic of Leiria

Shak Smith
Associate Researcher, Computer Science & Communication Research Centre

School of Management and Technology

Department of Computer Engineering

Master in Cybersecurity & Digital Forensics

Leiria, August 2024

This page intentionally left blank.

Declaration of Authorship

Has undersigned, hereby it his declared that this work entitled "Improving Machine Learning Efficiency Against Noisy Data Sources" is the original work and that it has not previously in its entirety or in part been submitted at any university or higher education institution for the award of any degree, diploma, or other qualifications. It is also hereby declared that to the best of the knowledge, this work contains no material previously published or written by another person, except where due reference, acknowledgement, and citation is made.

Leiria, August 2024

—————————————————————

Joe Smith

Acknowledgements

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Resumo

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Palavras-Chave: Keyword A, Keyword B, Keyword C.

ABSTRACT

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Keywords: Keyword A, Keyword B, Keyword C.

This page intentionally left blank.

CONTENTS

Appendices

This page intentionally left blank.

LIST OF FIGURES

List of Tables

Glossary

Latex   Typesetting system commonly used for the production of scientific and mathematical documents due to its powerful handling of formulas and references. It is widely used for the preparation of documents such as research papers, theses, reports, and articles. *(p. 5)*

Mathematics   Mathematics is what mathematicians do. *(p. 5)*

This page intentionally left blank.

Acronyms

This page intentionally left blank.

1

INTRODUCTION

Welcome to the introduction of your dissertation. The introduction of a dissertation serves as a critical component, setting the tone and laying the foundation for the entire research endeavour. It is tasked with providing a clear and concise overview of the research topic, elucidating the context and significance of the study within the broader academic landscape. A well-crafted dissertation introduction should delineate the research problem or question, offering a rationale for its relevance and addressing any existing gaps in knowledge. Furthermore, it typically outlines the objectives and aims of the study, guiding the reader through the anticipated contributions and outcomes. In addition, the introduction often encapsulates the methodology employed, presenting the chosen approach and rationale behind it. Lastly, it functions as a road-map, offering a brief glimpse into the structure and organisation of the dissertation, thereby orienting the reader and facilitating comprehension of the subsequent chapters. Overall, a dissertation introduction should engage the reader's interest, provide a clear framework for the research, and justify its importance in the academic realm. For a clearer and reader-friendly experience on referencing chapters, please refer to the chapter titled Citations & Other Elements (referred to as Chapter 4).

2

DOCUMENT VARIABLES

In this document, it is possible to change the value of various variables, such as the author's name, the title of the work, and others, so that the author has greater flexibility in their work. The file that operates these variables is `Variables/Variables.tex`. Below, in Table 2.1, you can see all the variables that can be changed, as well as whether or not they must be filled in for the theme to work correctly. Please note that in order to omit a variable, it is only necessary to comment it out. Additionally, if you wish to suggest any other variable, please contact me via the contacts available on GitHub or in the class file of the template.

Table 2.1: *Document variables that can be changed.*

Variable	Mandatory	Optional
Title	✓	-
Subtitle	✓	-
University	✓	-
School	✓	-
Department	✓	-
Degree	✓	-
Course	-	✓
Local and date	✓	-
First author name	✓	-
First author identification	✓	-
Second author name	-	✓
Second author identification	-	✓
Third author name	-	✓
Third author identification	-	✓
Supervisor name	✓	-
Supervisor e-mail	✓	-
Supervisor title and affiliation	✓	-
Co-supervisor name	-	✓
Co-supervisor e-mail	-	✓
Co-supervisor title and affiliation	-	✓
Second co-supervisor name	-	✓
Second co-supervisor e-mail	-	✓
Second co-supervisor title and affiliation	-	✓

3

Custom Notes

In the course of writing, it is sometimes necessary to insert notes in the middle of the document so that in the future we can return to them and read what we have written. Sometimes we need a space to jot down a quick `TODO`. With this in mind, two commands have been developed: `\todo{TEXT}` and `\note{TEXT}`. These commands accept normal text and produce two different types of boxes, respectively, for each type of use. These boxes are "breakable", which means that they can be placed anywhere in the document and will behave as if they were text. Below is a simple usage of both environments. Please note that both of these environments were created with the intention of helping you develop your work and **should not be used in the final version of your document!**

> **TODO**
>
> Write about X! Later, we should remove Y.

> **NOTE**
>
> Come to me, all you who are weary and burdened, and I will give you rest. Take my yoke upon you and learn from me, for I am gentle and humble in heart, and you will find rest for your souls.

4

CITATIONS & OTHER ELEMENTS

In this chapter, we provide detailed guidance on the correct procedures for citing and referencing various elements within your document. Specifically, we will cover the proper methods for citing chapters, referencing figures and tables. We also provide information on how you can cite external works provided by a BibTeX bibliography.

4.1 Citations

We present two distinct approaches for citing entries in the bibliography. The first method involves in-text citations, executed using `\citet{ENTRY}`, while the second method employs `\citep{ENTRY}` for citations within a paragraph. Below is an example demonstrating both usages. It's essential to note that you can cite multiple works within the same citation environment. To achieve this, you should use the following format: `\citep{ENTRY1, ENTRY2, ...}`. It is also possible to cite only the title of the work or the author of the same. To do this, please use `\citetitle{ENTRY}` for title citations and `\citeauthor{ENTRY}` for author citations.

> Proper citations play a crucial role in academic writing, serving as the foundation for credibility, transparency, and the advancement of knowledge. They are a fundamental aspect of responsible scholarly writing. Please ensure accurate and appropriate citations.

Example: A novel signature scheme is introduced, along with an implementation of the Diffie-Hellman key distribution scheme that accomplishes a public key cryptosystem (Elgamal, 1985). According to Elgamal (1985), a new signature scheme that accomplishes a public key cryptosystem is introduced.

4.2 References

Much like citations, it is advisable to employ references in your document for citing crucial elements such as chapters, sections, figures, or tables. To reference these ele-

ments, begin by creating a label. This label can be generated using `\label{TEXT}`, and it should be positioned within the element you intend to refer to. Once the element is created, you can utilise `\ref{LABEL}` to generate an in-text reference. **We strongly recommend using** `\autoref{LABEL}`. This command automatically creates a custom link with colour corresponding to the type of element being referred to. For instance, a chapter reference will appear like this: Chapter 1, rather than simply Chapter 1.

> Just as with citations, ensuring proper references to elements within the document is of paramount importance. Remember to reference chapters and sections when necessary, **and consistently refer to other elements such as figures, tables, or listings.**

4.3 Glossary & Acronyms

The document includes both a glossary and an acronym list, accessible at the beginning of the document. You can create a new entry in either the `Miscellaneous/02-Glossary` or `Miscellaneous/03-Acronyms` sections, depending on the type of entry you intend to add. Once the entry is created, you can reference it using `\gls{ENTRY}` for glossary entries. For acronym entries, there are two ways to reference them. The first method, `\acrfull{ENTRY}`, should be used the first time the acronym appears in the text as it automatically provides the definition in-text. Subsequently, to refer to the acronym without repeating its meaning, use `\acrshort{ENTRY}`.

Example: Utilising Latex for Mathematics is essential (...). It is advisable to seek both the Greatest Common Divisor (GCD) and Least Common Multiple (LCM) because (...). Subsequently, with the aid of GCD and LCM, we can (...).

5

FIGURES

In LaTeX, integrating figures is a straightforward process. To insert them, you should utilise the environment `\begin{figure}`. You can customise the `width` parameter according to your requirements, but it is crucial to select a high-quality figure when inserting it into your documents. It is equally crucial to furnish a well-crafted caption. If necessary, consider including citations or references to indicate the figure's origin. The caption environment is denoted as `\caption{TEXT}`. To generate a smaller caption for the Table of Figures, be sure to utilise the format `\caption[SMALL_TEXT]{BIG_TEXT}`. By following the aforementioned tips, we can create a figure as demonstrated in Figure 5.1.
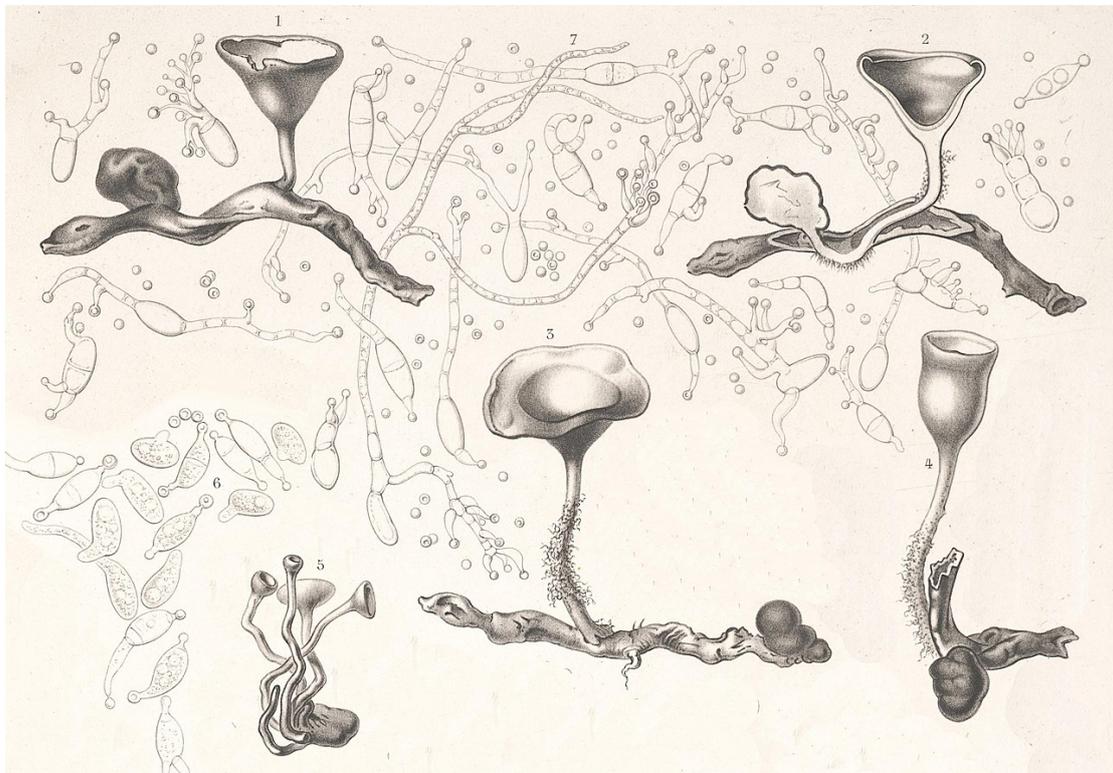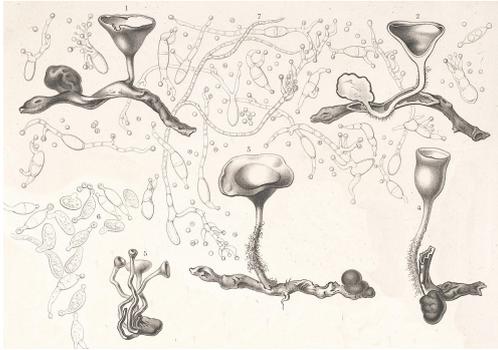

Figure 5.1: *Illustration of the fungus Dumontinia tuberosa by physician, mycologist, and illustrator Charles Tulasne (1816–1884) in the book Selecta Fungorum Carpologia (1861–65). (Name of the original work: Peziza tuberosa parasite on Anemone nemorosa).*

5.1 Side-by-Side Figures

For the purpose of comparing or for other reasons, you can insert side-by-side figures using both the \begin{figure} and \begin{subfigure} environments. You can also refer to the sub-figure as Figure 5.2a and Figure 5.2b.

(**a**) *Caption for figure 1.*   (**b**) *Caption for figure 2.*

Figure 5.2: *Overall caption for the figure.*

To customise the spacing between sub-figures, utilise the \hspace{VALUE} command. Establishing adequate spacing is crucial for enhancing visual appeal and ensuring a reader-friendly experience. Below is a code snippet that represents the Figure 5.2 - both label and caption text were omitted.

```
\begin{figure}[!htpb]
    \centering
    \begin{subfigure}{0.45\textwidth}
        \includegraphics[width=\textwidth]{FIGURE_PATH}
        \caption{TEXT}
        \label{TEXT}
    \end{subfigure}
    \hspace{.5cm} % Adjust the space as needed
    \begin{subfigure}{0.45\textwidth}
        \includegraphics[width=\textwidth]{FIGURE_PATH}
        \caption{TEXT}
        \label{TEXT}
    \end{subfigure}
    \caption{TEXT}
    \label{TEXT}
\end{figure}
```

6

TABLES

Tables play a vital role in presenting your findings effectively. In this chapter, we delve into various techniques for conveying information through tables, employing different environments available in this template. Although defining tables in LaTeX may appear complex, using this template makes the process more straightforward.

> Prior to showcasing the different table environments, it's crucial to note that each one must be enclosed within a `\begin{table}` environment. Additionally, it is recommended to utilise the `[!htpb]` float options for improved document placement. **This advice should be taken into consideration when positioning figures as well**.

6.1 Tabular Environment

The conventional `\begin{tabular}` environment enables you to create a simple yet elegant table. Table 6.1 is generated using a centering environment for added emphasis. It also incorporates the `booktab` configuration for a more sophisticated table style.

Table 6.1: *A table showcasing the usage of the tabular environment.*

Header 01	Header 02	Header 03
Lorem Ipsum	Lorem Ipsum	✓
Lorem Ipsum	Lorem Ipsum	✓
Lorem Ipsum	Lorem Ipsum	-
Lorem Ipsum	Lorem Ipsum	-
Lorem Ipsum	Lorem Ipsum	✓

6.2 Tabularx Environment

Employ the `\begin{tabularx}` package to construct a table featuring automatically expanding multi-columns. To achieve this automatic behaviour for multi-columns, utilise the following environment: `\begin{tabularx}{\textwidth}{@{}lX@{}}`. Take note that we substitute X in place of l or c, explicitly indicating that the column will

function as a multi-column, occupying the entire available space. Table 6.2 showcases the usage of the `begin{tabularx}` environment.

Table 6.2: *A table showcasing the usage of the tabularx environment.*

Header 01	Header 02
Foo Bar Baz	Quisque cursus, metus vitae pharetra auctor, sem massa mattis sem, at interdum magna augue eget diam.
Ipsum Dolor	Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Curabitur aliquet quam id dui.
Dolor Sit	Phasellus condimentum elementum justo, quis interdum est sagittis ac. Vestibulum non arcu sit amet justo lobortis semper.
Amet Consectetuer	Integer nec odio praesent libero sed cursus ante dapibus diam sed nisi vestibulum non arcu.
Consectetuer Adipiscing	Nulla quis sem at nibh elementum imperdiet. Duis sagittis ipsum. Praesent mauris.

6.3   Longtable Environment

At times, when dealing with exceptionally lengthy tables, it becomes necessary to split them across multiple pages. In LaTeX, this can be achieved using the `\begin{longtable}` environment. Feel free to consult Table 6.3 for a detailed demonstration of how the `longtable` operates.

Table 6.3: *A table showcasing the usage of the longtable environment.*

Names	E-Mails	Job/Role
Alice Johnson	alice.johnson@email.com	Project Manager
Bob Thompson	bob.thompson@email.com	Data Analyst
Charlie Davis	charlie.davis@email.com	Marketing Specialist
David Miller	david.miller@email.com	QA Tester
Emily White	emily.white@email.com	Graphic Designer
Frank Martin	frank.martin@email.com	HR Coordinator
Grace Turner	grace.turner@email.com	Financial Analyst
Henry Lee	henry.lee@email.com	System Administrator
Ivy Carter	ivy.carter@email.com	Customer Support
Jack Wilson	jack.wilson@email.com	Frontend Developer
Jane Reed	jane.reed@email.com	UX Designer
Kevin Evans	kevin.evans@email.com	Product Manager
Linda Adams	linda.adams@email.com	Accountant
Mike Hill	mike.hill@email.com	Network Engineer
Nina Garcia	nina.garcia@email.com	Business Analyst
Oliver Smith	oliver.smith@email.com	Sales Representative
Pamela Turner	pamela.turner@email.com	Legal Counsel
Quincy Brown	quincy.brown@email.com	IT Consultant

Table 6.3 continued from previous page

Names	E-Mails	Job/Role
Rachel Moore	rachel.moore@email.com	Content Writer
Samuel White	samuel.white@email.com	Research Scientist

6.4   Complex Tables

Creating intricate tables in LaTeX can be a somewhat challenging task. Therefore, we highly recommend using the Table Generator. With this tool, you can design your table with the desired style and then easily copy and paste it into your document. This approach simplifies the process and helps ensure the accurate representation of complex tables in your LaTeX document. However, it's crucial to keep in mind that a table should be easily comprehensible for the reader and should not be overly complex. The complexity of a table may impede understanding. For example, Table 6.4 presents a table with intricate details.

Table 6.4: *A table showcasing the usage of the complex tables.*

Category	Details	
	Subcategory	Carried Out
Long Category Name A	Long Subcategory Name A	✓
	Ipsum	✓
	Adipiscing	-
Long Category Name B	Long Subcategory Name B	-
	Ipsum	-
	Adipiscing	-
Long Category Name C	Long Subcategory Name C	✓
	Consectetur	✓
	Adipiscing	-

7

LISTS

Creating lists in LaTeX is straightforward, offering various options to suit your needs. You can generate a bullet list using \begin{itemize}, or opt for a numbered list with \begin{enumerate}. Below is an example with the \begin{itemize} environment.

- List entries start with the \item command.
- Individual entries are indicated with a black dot, a so-called bullet.
- The text in the entries may be of any length.

As mentioned earlier, you can generate a numbered list using the \begin{enumerate} environment. Here is an example:

1. Items are numbered automatically.
2. The numbers start at 1 with each use of the enumerate environment.
3. Another entry in the list.

You can also nest list entries by creating a list inside another list of the same type. Here is an example:

1. First level item
2. First level item
   (a) Second level item
   (b) Second level item
       i. Third level item
       ii. Third level item
           A. Fourth level item
           B. Fourth level item

> Please note that the labels change automatically regardless of the environment being the same for every list. **This demonstrates that there's no need to worry about changing the environment for something different.** However, if desired, you have the flexibility to do so.

You can also modify the label of your list to something entirely different that suits your needs. To accomplish this, insert a new `\item` and enclose your desired label in square brackets. For example, `\item[!]` will result in an exclamation point as your new label. Below are some examples of modified labels.

- This is my first point
- Another point I want to make
! A point to exclaim something!
■ Make the point fair and square.
  A blank label?

Finally, you can create a description list. Unlike having a bullet point or a numbered label, a description list enables you to use custom descriptions that suit your list. In the example below, there are three `\item` entries: one without a label, and two with descriptions.

Item 1: This is the first item with a description.
Item 2: Another item with a different description.
  An item without a specific label.

8

CODE LISTINGS

At times, you may want to include source code from your programs and applications within your document. To achieve this, you can use two nested environments: \begin{listing} to create a listing with both caption and label, and \begin{minted} for code highlighting. Listing 8.1 provides an example of a source code in C.

```c
#include <stdio.h>
int main() {
    printf("Hello, World!"); /*printf() outputs the quoted string*/
    return 0;
}
```

Listing 8.1: *Hello world in C.*

The code mentioned above was inserted into the document. However, an alternative approach is to input your code from an external file. To do so, you just need to use the command \inputminted{CODE_LANGUAGE}{FILE}. Of course, you should place that command inside of the \begin{listing} environment. Listing 8.2 illustrates an example of Octave source code that has been input from an external file.

```octave
% Function to compute the sum without charge of two vectors
function X = BitXorMatrix(A,B)
    % Convert elements into usigned integers
    A = uint8(A);
    B = uint8(B);

    m1 = length(A);
    m2 = length(B);
    X = uint8(zeros(m1, m2));
    for n1=1:m1
        for n2=1:m2
            X(n1, n2) = bitxor(A(n1), B(n2));
        end
    end
```

Listing 8.2: *XOR operation in Octave.*

In some cases, when you simply want to highlight a specific command, it's recommended not to use `listing` or `minted`. Instead, you should utilise the `\verb` command for inline highlighting or the `\begin{verbatim}` environment for longer sections of highlighted code. An example of a lengthy `verbatim` section is provided below, demonstrating how to create a `listing` with an input code:

```
\begin{listing}[!htpb]
    \inputminted{CODE_LANGUAGE}{FILE}
    \caption{TEXT}
    \label{TEXT}
\end{listing}
```

Sometimes it is necessary to display longer code that occupies more than one page. For this purpose, please use the environment `\begin{longlisting}`. This environment will easily break your code into multiple pages for better readability without you worrying about the size of your code. An example is shown below in Listing 8.3.

```lisp
 1  (defun factorial (n)
 2    "Calculate the factorial of a number."
 3    (if (zerop n)
 4        (* n (factorial (1- n)))))
 5
 6  (defun fibonacci (n)
 7    "Calculate the nth Fibonacci number."
 8    (cond ((zerop n) 0)
 9          ((= n 1) 1)
10          (t (+ (fibonacci (1- n)) (fibonacci (- n 2))))))
11
12  (defun gcd (a b)
13    "Calculate the greatest common divisor of a and b."
14    (if (zerop b)
15        a
16        (gcd b (mod a b))))
17
18  (defun primes-up-to (limit)
19    "Return a list of all prime numbers up to LIMIT."
20    (let ((primes '()))
21      (loop for i from 2 to limit
22            unless (some (lambda (p) (zerop (mod i p))) primes)
23            do (push i primes))
24      (nreverse primes)))
25
26  (defun example-function (x)
27    "An example function to demonstrate Lisp capabilities."
28    (let ((result (list (factorial x)
29                        (fibonacci x)
30                        (gcd x 10)
31                        (primes-up-to x))))
```

```lisp
32      (format t "Factorial of ~A: ~A~%" x (factorial x))
33      (format t "Fibonacci of ~A: ~A~%" x (fibonacci x))
34      (format t "GCD of ~A and 10: ~A~%" x (gcd x 10))
35      (format t "Primes up to ~A: ~A~%" x (primes-up-to x))
36    result))
37
38 (example-function 10)
```

Listing 8.3: *A sample of functions in Lisp.*

9

CONCLUSION

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Bibliography

Elgamal, T. (July 1985). "A public key cryptosystem and a signature scheme based on discrete logarithms". In: *IEEE Transactions on Information Theory* 31.4, pp. 469–472. ISSN: 1557-9654. DOI: 10.1109/TIT.1985.1057074. URL: https://ieeexplore.ieee.org/document/1057074.

Appendices

A

Appendix A

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

B

Appendix B

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

ANNEXES

M

Annex A

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.