



Lehrstuhl Angewandte Informatik IV
Datenbanken und Informationssysteme
Prof. Dr.-Ing. Stefan Jablonski

Institut für Angewandte Informatik
Fakultät für Mathematik, Physik und Informatik
Universität Bayreuth

Bachelorarbeit

Klaus Freiberger

Juni 1, 2019

Version: Final

Universität Bayreuth

Fakultät Mathematik, Physik, Informatik

Institut für Informatik

Lehrstuhl für Angewandte Informatik IV

Rapid Miner NLP Tools: Syntax Parsing

Bachelorarbeit

Klaus Freiberger

- | | |
|--------------------|---|
| <i>1. Reviewer</i> | Prof. Dr.-Ing. Stefan Jablonski
Fakultät Mathematik, Physik, Informatik
Universität Bayreuth |
| <i>2. Reviewer</i> | Dr. Lars Ackermann
Fakultät Mathematik, Physik, Informatik
Universität Bayreuth |
| <i>Supervisors</i> | Stefan Jablonski and Lars Ackermann |

Juni 1, 2019

Klaus Freiburger

Bachelorarbeit

Rapid Miner NLP Tools: Syntax Parsing, Juni 1, 2019

Reviewers: Prof. Dr.-Ing. Stefan Jablonski and Dr. Lars Ackermann

Supervisors: Stefan Jablonski and Lars Ackermann

Universität Bayreuth

Lehrstuhl für Angewandte Informatik IV

Institut für Informatik

Fakultät Mathematik, Physik, Informatik

Universitätsstrasse 30

95447 Bayreuth

Germany

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.

Abstract (different 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.

Contents

1	Introduction	1
1.1	Postcards: My Address	2
1.2	Motivation and Problem Statement	2
1.3	Results	3
1.3.1	Some References	3
1.4	Thesis Structure	3
2	Natural Language Processing	7
2.1	Grundlagen	7
2.2	Conclusion	8

Introduction

” *You can’t do better design with a computer, but you can speed up your work enormously.*

— **Wim Crouwel**

(Graphic designer and typographer)

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.

This is the second paragraph. 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.

1.1 Postcards: My Address

Ricardo Langner

Alfred-Schrapel-Str. 7

01307 Dresden

Germany

1.2 Motivation and Problem Statement

And after the second paragraph follows the third paragraph. 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.

After this fourth paragraph, we start a new paragraph sequence. 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.

[Jurgens:2000; Jurgens:1995; Miede:2011; Kohm:2011; Apple:keynote:2010; Apple:numbers:2010; Apple:pages:2010]

1.3 Results

This is the second paragraph. 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.

1.3.1 Some References

[WEB:GNU:GPL:2010; WEB:Miede:2011]

1.4 Thesis Structure

Chapter ??

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.

Chapter ??

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.

Chapter ??

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.

Chapter ??

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.

Chapter ??

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.

Natural Language Processing

“ *A picture is worth a thousand words. An interface is worth a thousand pictures.*

— **Ben Shneiderman**
(Professor for Computer Science)

2.1 Grundlagen

Betrachtet man einen Satz in einer natürlichen Sprache, die im Rahmen dieser Arbeit auf Englisch festgelegt ist, so kann ein Computer dessen Inhalt nicht ohne Weiteres herauslesen. Hierfür bedarf es verschiedener Hilfsstrukturen und zusätzlicher Informationen. Als ersten Schritt bietet es sich an, den einzelnen Wörtern eines Satzes ihre Wortart zuzuordnen. Mit Wortart, alternativ auch Wortklasse oder zu englisch part-of-speech (POS), ist gemeint, wie ein Wort im Satz auftritt. Beispiele für Wortarten sind Nomen, Verb und Adjektiv. In dieser Arbeit wird das Tagset, also die Menge an Wortarten, aus der Penn Treebank verwendet. Der Satz

My dog also likes eating sausage.

würde mit diesem Tagset also folgendermaßen annotiert werden:

My/PRP\$ dog/NN also/RB likes/VBZ eating/VBG sausage/NN ./.

Wie ein Satz maschinell mit POS-Tags versehen werden kann wird in dieser Arbeit allerdings nicht weiter behandelt. Über diese zusätzliche Notation hinaus kann man erkennen, dass sich in der englischen Sprache oftmals mehrere Wörter als Gruppe oder als eine Komponente innerhalb des Satzes verhalten. So eine Gruppe wäre zum Beispiel die Nominalphrase *My Dog* oder die Verbalphrase *likes eating sausage*. Auch wird wieder die Annotation der Penn Treebank verwendet. Der Satz welcher sich aus der zusätzlichen Annotation ergibt lautet:

(S
(NP (PRP\$ My) (NN dog))

(ADVP (RB also))
(VP (VBZ likes)
(S
(VP (VBG eating)
(NP (NN sausage))))))
(. .))

2.2 Conclusion

List of Figures

List of Tables

Declaration

You can put your declaration here, to declare that you have completed your work solely and only with the help of the references you mentioned.

Bayreuth, Juni 1, 2019

Klaus Freiburger

