



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

Philipp Scholz

23.April 2019

Version: Draft

Universität Bayreuth

Fakultät Mathematik, Physik, Informatik

Institut für Informatik

Lehrstuhl für Angewandte Informatik IV

NLP-Plattform: Integration und Evaluation von POS-Tagging-Algorithmen

Bachelorarbeit

Philipp Scholz

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

23.April 2019

Philipp Scholz

Bachelorarbeit

NLP-Plattform: Integration und Evaluation von POS-Tagging-Algorithmen, 23.April 2019

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

Supervisors: Lars Ackermann and Stefan Jablonski

Universität Bayreuth

Lehrstuhl für Angewandte Informatik IV

Institut für Informatik

Fakultät Mathematik, Physik, Informatik

Universitätsstrasse 30

95447 Bayreuth

Germany

Abstrakt

Im Themenbereich *Natural Language Processing* (kurz NLP) versucht die Informatik, natürliche Sprachen für Algorithmen zugänglich und interpretierbar zu machen. Ein wichtiger Zwischenschritt ist die Identifizierung der syntaktischen Bedeutung von Wörtern (*Parts of Speech*, kurz POS) in gesprochener Sprache, und deren Zuweisung in Form von Tags (POS-Tagging). Für diese Aufgabe existiert eine Vielzahl unterschiedlich leistungsfähiger und robuster Algorithmen.

Im Rahmen dieser Arbeit wurde eine Sammlung solcher POS-Tagging-Algorithmen zusammen mit einem Evaluationssystem als Operatoren im Programm RapidMiner (zur Verfügung gestellt von RapidMiner GmbH) implementiert.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung dieser Arbeit	1
1.2	Part-of-Speech-Tagging allgemein	2
1.2.1	Tagset	3
1.2.2	Korpus und Treebank	3
1.2.3	Evaluationsmetriken	3
1.3	Aufbau der Arbeit	5
2	Verwandte Arbeiten	7
2.1	Stanford Tagger	7
2.1.1	Tokenization-Unterschiede	7
3	Konzept	9
3.1	Probleme und Ziele	9
3.1.1	Tokenization-Unterschiede	9
4	Implementierung	11
4.1	Planung	11
4.1.1	Inputformat	11
4.1.2	Format für Tagger-Ergebnisse	12
4.2	Struktur	12
4.2.1	Übersicht	12
4.3	Evaluation	14
4.3.1	Parsing	14
4.4	Conclusion	14
5	Evaluation	15
5.1	Goldstandard-Korpus	15
5.2	Concepts Section 2	15
5.3	Concepts Section 3	15
5.4	Conclusion	15
6	Conclusion	17
6.1	System Section 1	17
6.2	System Section 2	18

6.3 Future Work	20
---------------------------	----

Einleitung

:TODO :NC

Keine bekannte Lebensform hat eine mit der menschlichen Sprache vergleichbar komplexe Form von Informationsaustausch entwickelt [Rao]. Von selbst ergibt sich die Frage, ob und wie Sprache maschinell verarbeitet werden kann, um sie unter anderem zu interpretieren oder zusammenzufassen. Antworten auf diese Problemstellung liefert das Teilgebiet der Informatik *Natural Language Processing* (Dt. Verarbeitung natürlicher Sprachen, kurz NLP). NLP gliedert sich in viele Teilbereiche: Strukturelle Analyse, Semantik, Phonetik und einige weitere. In dieser Arbeit konzentrieren wir uns auf einen wichtigen Bestandteil der strukturellen Analyse, dem korrekten Identifizieren von syntaktischen Rollen von Wörtern und Symbolen (*Parts of Speech*, kurz POS), bezeichnet als *POS-Tagging* [Smith].

Ein Algorithmus, der POS-Tagging betreibt (POS-Tagger), nimmt die Sprache in Textform an und gibt ihn üblicherweise mit Tags versehen wieder aus, wie beispielsweise der Text

I like the blue house.

zu folgendem verarbeitet wird:

I\PRONOUN like\VERB the\DET blue\ADJ house\NOUN .\.

Wie die tatsächliche Ausgabe eines Taggers exakt formatiert wird und welche Tags auftreten, wird später angesprochen. Wichtiger ist hingegen, dass POS-Tagger diese Tags nicht garantiert korrekt wählen (siehe Abschnitt 1.2). Es ist also von Interesse, deren Performance zu bewerten.

1.1 Problemstellung dieser Arbeit

Die Datenverarbeitungs-Plattform *RapidMiner* :NC bietet im Rahmen der Erweiterung *Text Processing* :NC Funktionen (*Operatoren*), um Texte einzulesen und zu verarbeiten. NLP ist in RapidMiners Textverarbeitungs-Erweiterungen jedoch weitgehend noch nicht implementiert. Ziel dieser Arbeit ist es, in Form einer auf *Text Processing*

aufbauenden Erweiterung sowohl POS-Tagger zu implementieren, als auch ein Evaluationsrahmenwerk, mit dem deren Performance gemessen werden kann.

:TODO ?

Die implementierte Erweiterung liefert :TODO[anzahl] Tagging-Operatoren, :TODO[anzahl] Evaluationsoperatoren, ein spezialisiertes Übergabeformat für Tagging-Ergebnisse und eine Standardisierung für verwendete Tagsets :TODO[genauer?]. Gleichzeitig sind die Operatoren in ihrem Ein- und Ausgangsformat weitgehend kompatibel mit der Erweiterung *Text Processing*.

1.2 Part-of-Speech-Tagging allgemein

Betrachtet man das Wort „like“ aus dem einleitenden Beispiel, dann fällt auf, dass es alternativ zum Verb „mögen“ auch als Präposition „wie“ interpretiert werden könnte, auch wenn intuitiv schnell erkennbar wird, dass letztere Variante falsch ist. Diese Uneindeutigkeit (*Ambiguität*) ist das zentrale zu lösende Problem für POS-Tagger [Smith]; Im Gegensatz zum Menschen kann ein Algorithmus Ambiguitäten nicht intuitiv auflösen. Aus diesem Grund arbeiten POS-Tagger nicht zwingend vollständig korrekt.

Während NLP viele andere Analyseaufgaben neben POS-Tagging zusammenfasst, sind diese bei moderneren und komplexeren NLP-Algorithmen nicht mehr strikt von POS-Tagging trennbar, wenn die Performance des Taggers maximiert werden soll :NC. Zusatzinformationen, die parallel erarbeitet werden können, wie z.B. die Satzstruktur (u.A. Identifizierung von Teilsätzen), können das auflösen von Ambiguitäten erheblich erleichtern. Zur Vereinfachung betrachten wir aber in dieser Arbeit nur POS-Tagging selbst, welches typischerweise in folgende zwei Arbeitsschritte unterteilt wird [Smith]:

Sequenzialisierung: Zuerst muss bestimmt werden, welche Teile des angegebenen Dokuments jeweils ein Tag erhalten. Hierzu wird jedes Wort und jedes zusammenhängende Satzzeichen als ein *Token* ausgegeben. Die Reihenfolge der Wörter bleibt als Reihenfolge der Token hierbei erhalten.

Tagging: Mit Hilfe von statistischen, linguistischen und rechnerischen Methoden wird jedem Token ein POS-Tag zugewiesen.

Es folgen noch ein paar weitere Erläuterungen und Definitionen:

1.2.1 Tagset

Um einheitliche Verarbeitung und Vergleichbarkeit zu ermöglichen, werden die Tags in *Tagsets* definiert, wie zum Beispiel dem des Penn-Treebank-Projekts [Web:PennBank] [Paper:PennBank].

1.2.2 Korpus und Treebank

Als *Korpus* bezeichnet man eine simple Ansammlung von Text. Der *WSJ-Korpus* :NC beispielsweise ist eine Sammlung von Artikeln der Nachrichtenartikel des *Wall Street Journal*.

Korpora, die zu Token-Ketten sequenzialisiert wurden und deren Token mit (POS-) Tags versehen wurden, werden als annotierte Korpora oder *Treebanks* bezeichnet. Zusätzlich ist eine Treebank mit anderen Informationen wie der Satzstruktur versehen, diese sind hier aber nicht weiter relevant. Treebanks sind in der Regel zu nahezu 100% korrekt annotiert. Ist dies der Fall, spricht man auch von einem *Goldstandard*. Da Goldstandards manuell korrigiert werden, können seltene Fehler auftreten. Diese sind allerdings zu wenige, um als relevant angesehen zu werden. :NC

1.2.3 Evaluationsmetriken

Um die Performance eines Taggers zu bewerten, müssen Metriken eingeführt werden. Diese entstehen aus Vergleichen zwischen annotierten Korpora und den Tagging-Ergebnissen. Es folgt eine kurze Auflistung gängiger Wege, die Ergebnisqualität eines Taggers zu berechnen:

Per-Tag-Accuracy

Die Per-Tag-Accuracy (ab hier nur *Accuracy*) ist die wichtigste und einfachste Form der Bewertung eines Taggers. Sie ist definiert als [Rao] :

$$\frac{\# \text{ korrekter Tags}}{\# \text{ aller Token}}$$

Per-Sentence-Accuracy

Analog zur Accuracy pro Tag kann auch Accuracy pro Satz definiert werden:

$$\frac{\# \text{ vollständig korrekter Sätze}}{\# \text{ aller Sätze}}$$

Hierzu müssen jedoch über Wort-Sequenzen hinaus auch Satz-Sequenzen definiert werden. In der Evaluation in Abschnitt 2.1 wurde dies ebenfalls berücksichtigt.
:TODO

Confusion Matrix und daraus resultierende Metriken

Konzentrieren wir uns auf bestimmte Tags (Labels), können wir genauere Aussagen Treffen (Für allgemeine Definitionen siehe [Rao], für mehrklassige Klassifizierungsprobleme siehe [Web:rxnlp]). Zuerst bilden wir die *Confusion Matrix*. In Abbildung 1.1 befindet sich ein Beispiel dafür. In dieser Matrix wird gegenübergestellt, wie viele Labels aus dem Goldstandard als welche Labels vom Tagger vorhergesagt wurden.

True Positives für Label X (kurz TP(X)) sind Vorhersagen, die dem Goldstandard entsprechen (in der Abbildung Gelb markiert).

False Positives (FP(X)) sind alle Vorhersagen X, die dem Goldstandard widersprechen.

True Negatives (TN(X)) sind alle Vorhersagen, die nicht X sind, und auch im Goldstandard nicht X lauten.

False Negatives (FN(X)) sind alle Vorhersagen, die nicht X sind, im Goldstandard aber X.

	GoldLabel_A	GoldLabel_B	GoldLabel_C	
Predicted_A	30	20	10	TotalPredicted_A=60
Predicted_B	50	60	10	TotalPredicted_B=120
Predicted_C	20	20	80	TotalPredicted_C=120
	TotalGoldLabel_A=100	TotalGoldLabel_B=100	TotalGoldLabel_C=100	

Abb. 1.1: Beispiel für eine Confusion Matrix für Labels A, B, C. Entnommen von [Web:rxnlp]

Daraus lassen sich Precision(Wie viele der Vorhersagen X waren korrekt?), Recall(Wie viele X im Goldstandard wurden korrekt erkannt?) und der F1-Score(Mittel aus Precision und Recall) berechnen:

$$Precision(X) = \frac{TP(X)}{TP(X) + FP(X)} = \frac{TP(X)}{\text{Vorhergesagte } X}$$

$$Recall(X) = \frac{TP(X)}{TP(X) + FN(X)} = \frac{TP(X)}{X \text{ im Goldstandard}}$$

$$F1-Score = 2 * \frac{Precision(X) * Recall(X)}{Precision(X) + Recall(X)}$$

Für alle drei dieser Metriken steht der Wert 1 für vollständig richtig und 0 für vollständig falsch.

1.3 Aufbau der Arbeit

Kapitel 2 betrachtet einige Implementierungen von Part-of-Speech-Taggern und deren Evaluation in anderen Projekten.

Kapitel 3 erläutert die Methodik und behandelten Probleme in diesem Projekt.

Kapitel 4 beschreibt detailliert die Implementierung und deren Zielsetzung der RapidMiner-Erweiterung.

Kapitel 5 führt Evaluationen der implementierten POS-Tagger anhand eines gewählten Goldstandards auf.

Verwandte Arbeiten

In diesem Kapitel soll auf einige Arbeiten verwiesen werden, in denen Evaluationen von vorgestellten POS-Taggern stattfinden. Es wird besonders beachtet, *wie* und nach welchen Metriken diese Ergebnisse berechnet werden.

2.1 Stanford Tagger

[Paper:StanfordTagger]

2.1.1 Tokenization-Unterschiede

Konzept

3.1 Probleme und Ziele

Da Tagging-Algorithmen sowohl nicht fehlerfrei arbeiten als auch in ihrem Ergebnisformat voneinander - und insbesondere von Goldstandards - abweichen, entstehen Unsicherheiten beim Arbeiten mit solchen Informationen. Welche Probleme besonders dominant sind und welche Zielsetzungen für das Projekt sich daraus ergeben, wird in den folgenden Abschnitten ausführlicher behandelt.

3.1.1 Tokenization-Unterschiede

Ein wichtiger Schritt zur Vorverarbeitung von Text ist es, ihn in *Tokens* zu zerlegen. Ein Token ist ein Abschnitt, zu dem ein Tag gehört, typischerweise einzelne Wörter oder Satzzeichen. Formal wird aus dem zusammenhängenden Wort beziehungsweise Text w eine Menge M von n Teilwörtern c_i

$$M = \{c_1, c_2, \dots, c_n\}$$

generiert. POS-Tagger wie der von NLP4J :NC und der Stanford University :NC übernehmen diese Zerlegung selbst. Nehmen wir nun an, M wäre die Zerlegung von w in einem Goldstandard, und ein Tagger produziert ein Ergebnis mit der Zerlegung:

$$M_{tag} = \{d_1, d_2, \dots, d_m\}$$

Hierbei sind d_i wieder Teilworte und m die Anzahl dieser. Spätestens wenn nun $n \neq m$ gilt, wird klar, dass M und M_{tag} nicht mehr einfach iterativ verglichen werden können, da ab der Stelle p , wo der Text unterschiedlich geteilt wurde, $c_i \neq d_i$ sein kann, wobei ($p \leq i \leq \min\{n, m\}$). Es ist sogar nicht auszuschließen, dass solche Fehler bereits vorher passieren, obwohl weiterhin $n = m$ gilt.

Eines der Ziele bei der Implementierung ist also, den Vergleich von zwei Token-Mengen robust gegen solche Fehler zu machen.

Implementierung

Dieses Kapitel behandelt die Zielsetzungen, konkrete Planung und die Implementierung der RapidMiner-Erweiterung.

4.1 Planung

Zuerst müssen wir betrachten, welche Begebenheiten die RapidMiner-Plattform uns bietet, d.h. welche Datentypen ein Parser annehmen muss, wie man Ergebnisse möglichst ohne Informationsverlust übergibt und Goldstandards richtig einliest.

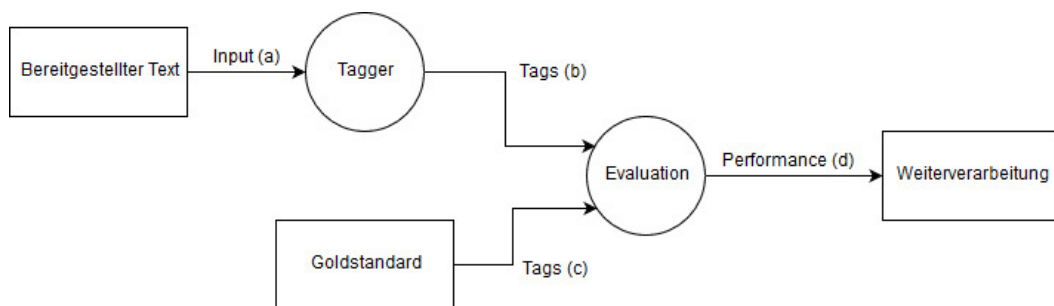


Abb. 4.1: Datenfluss in einem typischen Tagging- und Evaluationsprozess

Eine Übersicht hierzu bietet Abbildung 4.1, wobei insbesondere der Input (a) und Output (b) des Taggers interessant sind. Das Format des Goldstandards (c) ist exogen vorgegeben und das Ergebnis des Evaluationsprozesses muss lediglich RapidMiner-konform kodiert werden.

4.1.1 Inputformat

Für die Eingabe an den Tagger wäre es optimal, wenn Outputs anderer Textverarbeitender RapidMiner-Erweiterungen, insbesondere vom *Text Processing Plugin* :NC angenommen werden können. Hierzu kann einfach das Übergabeformat *Document* aus diesem Plugin verwendet werden.

4.1.2 Format für Tagger-Ergebnisse

Abhängig vom Tagger können Ergebnisse sehr Informationsreich sein. Der LingPipe-POS-Tagger liefert beispielsweise pro Token (Wort oder Symbol) nicht nur nicht nur das aus seiner Sicht plausibelste Tag (*First-Best*), sondern auch $n-1$ weitere, nachstehende Optionen (*N-Best*), wobei n gewählt werden kann :NC . Eine solche Menge an Metainformationen kann der Standard-Datentyp *Document* seitens RapidMiner nur schwer Kodieren. Eine neu entworfene Datenstruktur, die von Operator zu Operator übergeben werden können, bietet neue Freiheiten:

Segmentierung der Tag- und Token-Kette anhand von bestimmten Tags, die immer richtig erkannt werden, macht robuster gegen Verschiebungen im Ergebnis wie in Abschnitt 3.1.1 beschrieben: Wird Segment-weise verglichen und entstehen Verschiebungen durch ungleiche Token-Zerlegungen, dann verschwindet dieser Fehler, sobald ein neues Segment betreten wird, da die neuen betrachteten Segmente wieder an gleicher Stelle beginnen.

Anreicherung mit Zusatzinformationen ist möglich, da im Gegensatz zu einer einfachen Textuellen Darstellung auf ein Token nicht exakt ein Tag folgen muss. Hier kann beispielsweise pro Tag eine beliebig lange Liste der *N-Best* Tags stehen.

Ein solches alternatives Format ist jedoch von Operatoren, die nicht speziell dafür vorbereitet sind (wie der Evaluationsoperator), nicht lesbar. Darum müssen Ergebnisse auch immer als Typ *Document* angegeben werden.

4.2 Struktur

4.2.1 Übersicht

Abbildung 4.2 zeigt einen Überblick über die implementierten Klassen. Um das Diagramm überschaubar zu halten, wurden nur die Klassen eingezeichnet, die zusätzlich zum Erweiterungs-Template von RapidMiner :NC implementiert wurden. Außerdem wurden nur die notwendigen Enumerationen und Klassen für den NLP4J-Tagger aufgenommen, da andere Tagger analog strukturiert sind. Es folgt eine kurze Erklärung der verschiedenen Komponenten:

Operatoren: *Evaluator* und *Nlp4j_tagger* sowie alle anderen Klassen, die von der RapidMiner-Klasse *Operator* erben, sind auch in RapidMiner als Operatoren vertreten. Alle Operatoren außer *Evaluator* beinhalten einen POS-Tagger und überbrücken

4.3 Evaluation

Um Ergebnisse hinsichtlich ihrer Qualität zu prüfen, muss mit Goldstandards wie z.B. dem WSJ-Corpus :NC verglichen werden. Der Evaluationsoperator muss die Formate dieser Standards annehmen und interpretieren können. Neben dem eigentlichen Evaluieren ist also auch das *Parsen* eingehender Texte eine wichtige Aufgabe des Operators.

4.3.1 Parsing

Verschiedene Formate kodieren POS-Tags unterschiedlich. Eine einfache Notation ist die bereits bekannte Variante, bei der nach jedem Token ein „\“ und dann das Tag folgt. Andere Notationen kodieren zusätzlich Satzstrukturinformationen, sind also mittels Klammern geschachtelt, um z.B. Teilsätze zu markieren. Hier gibt es kein Symbol, das eindeutig ein POS-Tag ankündigt. Um die Parser-Methode des Evaluationsoperators präziser arbeiten können zu lassen, wurde ein Parameter hinzugefügt, in dem man das Format wählen kann. Es wurden zwei Modi für den Parser implementiert:

Backslash-Notation: In der oben genannten Notation, in der ausschließlich POS via „\“ markiert werden, ist Parsing einheitlich. Der Text wird an jedem Leerzeichen gespalten und jeder entstehende Substring am „\“. Das POS-Tag findet sich dann im zweiten String, der bei der zweiten Spaltung entsteht.

"None": Falls die Notation unbekannt ist, versucht der Parser, den Text an allen sinnvollen Symbolen, insbesondere dem Leerzeichen zu Trennen. Alle Substrings werden dann überprüft, ob sie ein POS-Tag sind.

Zusätzlich kann die Option „Ignore Brackets“ gewählt werden, falls Klammern für die Formatierung des einzulesenden Textes verwendet wurden. diese werden dann ignoriert.

Als Ergebnis gibt der Parser das Format TagString aus, das dann weiter verwendet werden kann. Der Tagset-Typ des TagStrings muss via Parameter angegeben werden.

4.4 Conclusion

Evaluation

Die folgenden POS-Tagger wurden implementiert:

Tagger (Training Korpus)	Performance (Korpus)
NLP4J	:TODO
LingPipe	:TODO
FastTag	-

Tab. 5.1: Liste der Implementierten Tagger. Performance laut eigenen Angaben.
* : Performance aus :NC

Die Performance in Tab. 5.1 bezieht sich auf eigene Angaben der Entwickler. Mit dem implementierten Evaluationsoperator sind wir allerdings in der Lage, die Tagger an einem eigenen annotierten Goldstandard-Korpus zu testen. Die folgenden Abschnitte diskutieren die Auswahl eines passenden Korpus und die Performance der Tagger auf diesem.

5.1 Goldstandard-Korpus

Ein annotierter Korpus muss dem Tagset der einzelnen Tagger entsprechen. Dies erweist sich als schwierig, da jeder Tagger u.U. Zeichen in unterschiedliche Token-Ketten spaltet und anders Kodiert. Es macht also Sinn, den Korpus und seine Tags auf den Tagger anzupassen, sofern das möglich ist, ohne den Informationsgehalt des Korpus zu ändern.

5.2 Concepts Section 2

5.3 Concepts Section 3

5.4 Conclusion

Conclusion

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Das hier ist der zweite Absatz. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

6.1 System Section 1

Und nun folgt – ob man es glaubt oder nicht – der dritte Absatz. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie

„Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Nach diesem vierten Absatz beginnen wir eine neue Zählung. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

6.2 System Section 2

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander

stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Das hier ist der zweite Absatz. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Und nun folgt – ob man es glaubt oder nicht – der dritte Absatz. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst

viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

6.3 Future Work

Nach diesem vierten Absatz beginnen wir eine neue Zählung. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“?

Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln. Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Abbildungsverzeichnis

1.1	Beispiel für eine Confusion Matrix für Labels A, B, C. Entnommen von [Web:rxnlp]	4
4.1	Datenfluss in einem typischen Tagging- und Evaluationsprozess	11
4.2	gekürztes Klassendiagramm	13

Tabellenverzeichnis

5.1	Liste der Implementierten Tagger. Performance laut eigenen Angaben. *	
	: Performance aus :NC	15

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst und dabei keine anderen als die angegebenen Hilfsmittel benutzt habe. Sämtliche Stellen der Arbeit, die im Wortlaut oder dem Sinn nach Publikationen oder Vorträgen anderer Autoren entnommen sind, habe ich als solche kenntlich gemacht. Die Arbeit wurde bisher weder gesamt noch in Teilen einer anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Bayreuth, 23.April 2019

Philipp Scholz

