

Fakultät Informatik

# My very fancy thesis title.

Bachelorarbeit im Studiengang Informatik

vorgelegt von

Lukas Welsch

Matrikelnummer 3201095

Erstgutachter:	Prof. Dr. Korbinian Riedhammer
Zweitgutachter:	Prof. Dr. Jens Albrecht
Betreuer:	Dipl. Ing. Andreas Sachs
Unternehmen:	Sopra Financial Technology GmbH

© 2020

Dieses Werk einschließlich seiner Teile ist **urheberrechtlich geschützt**. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Autors unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen sowie die Einspeicherung und Verarbeitung in elektronischen Systemen.



## Prüfungsrechtliche Erklärung der/des Studierenden

Angaben des bzw. der Studierenden:

Name: Welsch

Vorname: Lukas


Matrikel-Nr.: 3201095

Fakultät: Informatik



Studiengang: Informatik



Semester: Wintersemester  2020/21

### Titel der Abschlussarbeit:

Analyse und Visualisierung der Datenqualität innerhalb eines Data Warehouse

Ich versichere, dass ich die Arbeit selbständig verfasst, nicht anderweitig für Prüfungszwecke vorgelegt, alle benutzten Quellen und Hilfsmittel angegeben sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.

Muss noch ausgefüllt

Ort, Datum, Unterschrift Studierende/Studierender

## Erklärung zur Veröffentlichung der vorstehend bezeichneten Abschlussarbeit

Die Entscheidung über die vollständige oder auszugsweise Veröffentlichung der Abschlussarbeit liegt grundsätzlich erst einmal allein in der Zuständigkeit der/des studentischen Verfasserin/Verfassers. Nach dem Urheberrechtsgesetz (UrhG) erwirbt die Verfasserin/der Verfasser einer Abschlussarbeit mit Anfertigung ihrer/seiner Arbeit das alleinige Urheberrecht und grundsätzlich auch die hieraus resultierenden Nutzungsrechte wie z.B. Erstveröffentlichung (§ 12 UrhG), Verbreitung (§ 17 UrhG), Vervielfältigung (§ 16 UrhG), Online-Nutzung usw., also alle Rechte, die die nicht-kommerzielle oder kommerzielle Verwertung betreffen.

Die Hochschule und deren Beschäftigte werden Abschlussarbeiten oder Teile davon nicht ohne Zustimmung der/des studentischen Verfasserin/Verfassers veröffentlichen, insbesondere nicht öffentlich zugänglich in die Bibliothek der Hochschule einstellen.

Hiermit ☒ genehmige ich, wenn und soweit keine entgegenstehenden Vereinbarungen mit Dritten getroffen worden sind,  
☐ genehmige ich nicht,

dass die oben genannte Abschlussarbeit durch die Technische Hochschule Nürnberg Georg Simon Ohm, ggf. nach Ablauf einer mittels eines auf der Abschlussarbeit aufgebrachten Sperrvermerks kenntlich gemachten Sperrfrist

von Jahren (0 - 5 Jahren ab Datum der Abgabe der Arbeit),

der Öffentlichkeit zugänglich gemacht wird. Im Falle der Genehmigung erfolgt diese unwiderruflich; hierzu wird der Abschlussarbeit ein Exemplar im digitalisierten PDF-Format auf einem Datenträger beigelegt. Bestimmungen der jeweils geltenden Studien- und Prüfungsordnung über Art und Umfang der im Rahmen der Arbeit abzugebenden Exemplare und Materialien werden hierdurch nicht berührt.

Ort, Datum, Unterschrift Studierende/Studierender

Formular drucken

**Datenschutz:** Die Antragstellung ist regelmäßig mit der Speicherung und Verarbeitung der von Ihnen mitgeteilten Daten durch die Technische Hochschule Nürnberg Georg Simon Ohm verbunden. Weitere Informationen zum Umgang der Technischen Hochschule Nürnberg mit Ihren personenbezogenen Daten sind unter nachfolgendem Link abrufbar: <https://www.th-nuernberg.de/datenschutz/>



## Kurzdarstellung

Kurze Zusammenfassung der Arbeit, höchstens halbe Seite. Deutsche Fassung auch nötig, wenn die Arbeit auf Englisch angefertigt wird.

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.

## Abstract

*Only if thesis is written in English.*

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.



# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
1.1. This is an Important Section	1
1.1.1. And an even more important subsection	1
<b>2. Grundlagen</b>	<b>3</b>
2.1. Definition von Datenqualität	3
2.1.1. Datenqualitätsdimensionen	3
2.1.2. Definition / Metriken der Dimensionen	3
<b>3. Daten</b>	<b>5</b>
<b>4. Methoden</b>	<b>7</b>
4.1. Metriken	7
4.2. Bestehende Verfahren	7
4.3. Stakeholderanalyse	7
<b>5. Experimente</b>	<b>9</b>
<b>6. Ausblick</b>	<b>11</b>
<b>7. Fazit</b>	<b>13</b>
<b>A. Supplemental Information</b>	<b>15</b>
<b>Abbildungsverzeichnis</b>	<b>17</b>
<b>Tabellenverzeichnis</b>	<b>19</b>
<b>Auflistung</b>	<b>21</b>
<b>Literaturverzeichnis</b>	<b>23</b>
<b>Glossar</b>	<b>25</b>





# Kapitel 1.

## Einleitung

You can also write footnotes.<sup>1</sup>

### 1.1. This is an Important Section

It is possible to reference glossary entries as library as an example.

#### 1.1.1. And an even more important subsection

---

<sup>1</sup>Footnotes will be positioned automatically.



## Kapitel 2.

### Grundlagen

#### 2.1. Definition von Datenqualität

Datenqualität wird in der Literatur auf verschiedene Arten definiert.

##### 2.1.1. Datenqualitätsdimensionen

In der Literatur werden als Basisset der Datenqualitätsdimensionen Richtigkeit, Vollständigkeit, Konsistenz und Aktualität definiert. [[Scan 02](#)]

##### 2.1.2. Definition / Metriken der Dimensionen

Um diese eher abstrakten Definitionen messbar zu machen haben sich einige Verfahren etabliert. Diese lassen sich grob in die Kategorien subjektiv und objektiv aufteilen. [[Pipi 02](#)]  
Bei den subjektiven Verfahren werden die Grenzwerte, ab wann die Datenqualität schlecht ist durch Experten geschätzt. Die Experten geben oft auch ein Gefühl ab, ob sie die Datenqualität für gut oder schlecht halten. Der zweite Weg besteht darin die Datenqualität über mathematische Verfahren festzulegen. Allerdings werden auch hier häufig Experten benötigt, die die Grenzwerte schätzen.



## **Kapitel 3.**

### **Daten**

Die in der Arbeit verwendeten Daten stammen aus dem Data Warehouse einer deutschen Bank. Aufgrund des Datenschutzes wurden personenbezogene Daten pseudonymisiert und anonymisiert.



# Kapitel 4.

## Methoden

Um die nachfolgenden Methoden zur Analyse und Visualisierung von Datenqualität, zu vergleichen wurden sowohl qualitative als auch quantitative Untersuchungen durchgeführt.

### 4.1. Metriken

### 4.2. Bestehende Verfahren

### 4.3. Stakeholderanalyse

In this chapter, we're actually using some code!

```
1 x = 1
2 if x == 1:
3     # indented four spaces
4     print("x is 1.")
```

Listing 4.1: This is an example of inline listing

You can also include listings from a file directly:

```
1 x = 1
2 if x == 1:
3     # indented four spaces
4     print("x is 1.")
```

Listing 4.2: This is an example of included listing





## **Kapitel 5.**

### **Experimente**



## Kapitel 6.

### Ausblick



## Kapitel 7.

### Fazit



## Anhang A.

### Supplemental Information

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.

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.

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.



## Abbildungsverzeichnis



## **Tabellenverzeichnis**



## Auflistung

4.1. This is an example of inline listing . . . . .	7
4.2. This is an example of included listing . . . . .	7



## Literaturverzeichnis

- [Pip02] L. L. Pipino, Y. W. Lee, and R. Y. Wang. “Data Quality Assessment”. *Communications of the ACM*, Vol. 45, No. 4, pp. 211–218, Apr. 2002.
- [Scan02] M. Scannapieco and T. Catarci. “Data Quality under a Computer Science Perspective”. *Journal of The ACM - JACM*, Vol. 2, Jan. 2002.





# Glossar

**library** A suite of reusable code inside of a programming language for software development. i, 1

**shell** Terminal of a Linux/Unix system for entering commands. i