

The Extremes of Good and Evil

Master Thesis

presented by
Oliver Frendo (1510432),
Dandan Li (),
Zehui Wang () *and*
Yi-Ru Cheng ()

submitted to the
Data and Web Science Group
Prof. Dr. Right Name Here
University of Mannheim

November 25, 2015

Contents

1	Web Data Integration	1
1.1	Introduction to data and use case	2
1.2	Data translation	3
1.2.1	Data collection	3
1.2.2	Integrated schema	3
1.2.3	Data transformations	3
1.3	Identity resolution	4
1.3.1	Gold standards	4
1.3.2	Matching rules	4
1.3.3	Blocking functions	4
1.3.4	Learning matching rules	4
1.4	Data fusion	5
1.4.1	Input data	5
1.4.2	Gold standard	5
1.4.3	Conflict resolution functions	5
1.4.4	Accuracy	5

List of Figures

List of Tables

Chapter 1

Web Data Integration

1.1 Introduction to data and use case

1.2 Data translation

1.2.1 Data collection

Forbes

Freebase

DBpedia: Companies

DBpedia: Locations

1.2.2 Integrated schema

1.2.3 Data transformations

1.3 Identity resolution

1.3.1 Gold standards

Forbes \leftrightarrow Freebase

Freebase \leftrightarrow DBpedia companies

DBpedia companies \leftrightarrow DBpedia locations

1.3.2 Matching rules

1.3.3 Blocking functions

Notes on blocking function:

Freebase vs Forbes: - Country is always available in Forbes, sparse in Freebase (EXACT NUMBER)

Freebase vs DBPedia: - Country is somewhat sparse in both datasets (EXACT NUMBER) - DateFounded somewhat sparse in both (EXACT NUMBER)

Company vs Location - Country is only way to go

Own idea: Implement own partitioning blocker, but more fuzzy logic - I.E. Country first, then date, then location or so as alternative (EXACT NUMBER) - Generate more pairs, matching rule will take care if its same company or not- more conservative approach

1.3.4 Learning matching rules

1.4 Data fusion

1.4.1 Input data

1.4.2 Gold standard

1.4.3 Conflict resolution functions

1.4.4 Accuracy

Bibliography

Ehrenwörtliche Erklärung

Ich versichere, dass ich die beiliegende Master-/Bachelorarbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen. Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Mannheim, November 25, 2015

Unterschrift