

# VU Research Portal

## Extending Knowledge Bases from Human-readable Tables

Kruit, Benjamin Bienze

2021

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Kruit, B. B. (2021). *Extending Knowledge Bases from Human-readable Tables*.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Integrating Tables with Knowledge Bases . . . . .	14
1.2	Research Questions and Contributions . . . . .	16
1.2.1	Extracting New Facts . . . . .	16
1.2.2	Integrating N-ary Tables . . . . .	17
1.2.3	Creating a New KB from Tables . . . . .	18
1.2.4	Designing Pipelines . . . . .	19
1.3	Thesis Outline and Publications . . . . .	20
<b>2</b>	<b>Background and Related Work</b>	<b>23</b>
2.1	Tabular Data . . . . .	24
2.1.1	Data Modeling . . . . .	24
2.1.2	Web Tables . . . . .	28
2.1.3	Web Table Extraction . . . . .	29
2.1.4	Web Table Structures . . . . .	30
2.2	Knowledge Bases . . . . .	33
2.2.1	Overview of Knowledge Bases . . . . .	33
2.2.2	Representing N-ary Information . . . . .	36

2.2.3	Ontology Matching . . . . .	38
2.2.4	Knowledge Acquisition . . . . .	39
2.3	Table Interpretation . . . . .	41
2.3.1	Common Assumptions . . . . .	42
2.3.2	Approaches and Systems . . . . .	44
2.3.3	Constrained Table Interpretation . . . . .	46
2.4	Data Integration at Scale . . . . .	48
2.4.1	Schema Matching and Profiling . . . . .	48
2.4.2	Instance Matching and Entity Linking . . . . .	51
2.5	Impact and Applications . . . . .	52
2.5.1	Search, Query Answering and Integration . . . . .	53
2.5.2	Dataspaces . . . . .	54
2.6	Summary . . . . .	55
<b>3</b>	<b>Extracting Novel Facts from Tables</b>	<b>57</b>
3.1	Introduction . . . . .	58
3.2	Motivation: Measuring Redundancy . . . . .	59
3.2.1	Experiments . . . . .	61
3.3	Our Approach . . . . .	65
3.3.1	Background . . . . .	65
3.3.2	Intuition . . . . .	67
3.3.3	Table Interpretation . . . . .	69
3.3.4	Slot-Filling . . . . .	75
3.4	Evaluation . . . . .	76
3.4.1	Table Interpretation . . . . .	79
3.4.2	Measuring Redundancy . . . . .	81
3.4.3	Slot-filling . . . . .	83
3.5	Conclusion and Discussion . . . . .	84
<b>4</b>	<b>Extracting N-ary Facts from Table Clusters</b>	<b>87</b>
4.1	Introduction . . . . .	88

4.2	Background . . . . .	90
4.3	Method . . . . .	94
4.3.1	Reshaping . . . . .	95
4.3.2	Clustering . . . . .	101
4.3.3	Integration . . . . .	107
4.4	Evaluation . . . . .	109
4.4.1	Annotations . . . . .	110
4.4.2	Table Reshaping . . . . .	111
4.4.3	Table Clustering . . . . .	112
4.4.4	Table Interpretation and KB Integration . . . . .	113
4.5	Conclusion . . . . .	114
<b>5</b>	<b>Building a KB from Tables in Scientific Papers</b>	<b>117</b>
5.1	Introduction . . . . .	118
5.2	Background . . . . .	120
5.3	Overview . . . . .	123
5.4	Task 1: Table Interpretation . . . . .	126
5.4.1	Training Data Generation . . . . .	126
5.4.2	Table Header Detection . . . . .	128
5.4.3	Table Type Detection . . . . .	128
5.4.4	Column Type Detection . . . . .	129
5.5	Task 2: Entity Linking . . . . .	130
5.6	Evaluation . . . . .	134
5.6.1	Table Interpretation . . . . .	136
5.6.2	Entity Linking . . . . .	139
5.7	Conclusion . . . . .	140
<b>6</b>	<b>A Platform for Web Table Information Extraction</b>	<b>141</b>
6.1	Introduction . . . . .	142
6.2	Design of Takco . . . . .	143
6.2.1	Target Users . . . . .	146

6.3	Usage of Takco . . . . .	147
6.3.1	Usage as an Analysis Tool . . . . .	149
6.3.2	Usage as a Modular Platform . . . . .	151
6.4	Evaluation Results . . . . .	152
6.5	Conclusion . . . . .	154
<b>7</b>	<b>Conclusion</b>	<b>157</b>
7.1	Main Contributions and Future Work . . . . .	157
7.1.1	Extracting Novel Facts . . . . .	157
7.1.2	Integrating N-ary Tables . . . . .	160
7.1.3	Creating a New KB from Tables . . . . .	162
7.1.4	Designing Pipelines . . . . .	164
7.2	Final Reflections . . . . .	166
	<b>Bibliography</b>	<b>168</b>
	<b>Summary</b>	<b>191</b>
	<b>Curriculum Vitae</b>	<b>193</b>