On entity resolution in probabilistic data

Ayat, S.N.

Citation for published version (APA):

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
## Contents

1 Introduction ......................................... 1  
1.1 Uncertain Data .................................. 1  
1.2 Entity Resolution ................................. 2  
1.3 Entity Resolution for Probabilistic Data .............. 4  
1.4 Research Questions ................................ 5  
1.5 Roadmap .......................................... 7  

2 Background and Related Work ......................... 9  
2.1 Uncertain Data Models ............................ 9  
2.1.1 Possible Worlds Semantics .................... 9  
2.1.2 X-relation Data Model ....................... 11  
2.2 Entity Resolution ................................ 12  
2.2.1 Matching Methods ............................. 14  
2.2.2 Scalability Issues ............................ 24  
2.2.3 Entity Resolution for Probabilistic Data ........ 26  
2.2.4 Conclusion .................................. 29  

3 Entity Resolution for Probabilistic Data ............... 31  
3.1 Introduction .................................... 31  
3.2 Problem Definition ............................... 33  
3.2.1 Data Model ................................ 33  
3.2.2 Context-Free and Context-Sensitive Similarity Functions .... 33  
3.2.3 Problem Statement ......................... 34  
3.3 Context-Free Entity Resolution ................... 36  
3.3.1 CFA Algorithm ............................. 36  
3.3.2 Improving CFA Algorithm .................... 38  
3.4 Context-sensitive Entity Resolution ................. 41  
3.4.1 Monte Carlo Algorithm ....................... 42  
3.4.2 Parallel MC ................................. 42
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samenvatting</td>
<td>165</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>169</td>
</tr>
<tr>
<td>SIKS Dissertation Series</td>
<td>171</td>
</tr>
</tbody>
</table>