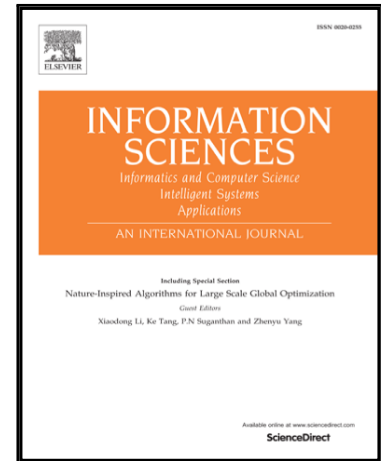


Accepted Manuscript

Experts Community Memory for Entity Similarity Functions Recommendation

Seung Hwan Ryu, Boualem Benatallah

PII: S0020-0255(16)31396-2  
DOI: [10.1016/j.ins.2016.10.045](https://doi.org/10.1016/j.ins.2016.10.045)  
Reference: INS 12604



To appear in: *Information Sciences*

Received date: 15 December 2015  
Revised date: 14 September 2016  
Accepted date: 22 October 2016

Please cite this article as: Seung Hwan Ryu, Boualem Benatallah, Experts Community Memory for Entity Similarity Functions Recommendation, *Information Sciences* (2016), doi: [10.1016/j.ins.2016.10.045](https://doi.org/10.1016/j.ins.2016.10.045)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Experts Community Memory for Entity Similarity Functions Recommendation

Seung Hwan Ryu\* and Boualem Benatallah

*School of Computer Science and Engineering, University of New South Wales, Sydney,  
NSW, 2051, Australia*

---

### Abstract

Similarity search (or similar entity search) is the process of finding all entities similar to a given entity (e.g., a person, a document, or an image). Although many techniques for similarity analysis have been proposed in the past, little work has been done on the question of which of the presented techniques are most suitable for a given similarity search task. Knowing the right similarity function is important as the task is highly domain- and data-dependent. In this article, we provide an approach for recommending which similarity functions (e.g., edit distance or jaccard similarity) should be used for measuring the similarity between two entities. The approach employs an incremental knowledge acquisition technique for capturing domain experts' knowledge about similarity functions and their usage contexts (e.g., entity class, attribute name and some keywords). In addition, for situations where domain experts have little or no knowledge about datasets, we analyze the features of the datasets and then suggest similarity functions based on the identified features. We also demonstrate the feasibility and effectiveness of our proposed approach on several real-world datasets from different domains.

*Keywords:* Entity matching, similarity function, incremental knowledge acquisition, similarity search, similarity measure, recommendation

---

---

\*Corresponding author

Download English Version:

<https://daneshyari.com/en/article/4944824>

Download Persian Version:

<https://daneshyari.com/article/4944824>

[Daneshyari.com](https://daneshyari.com)