



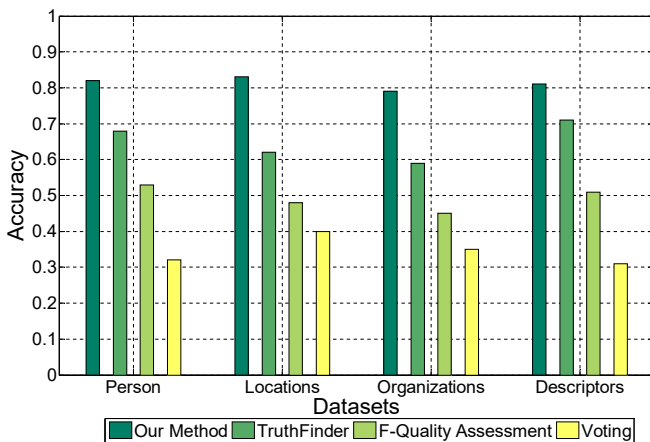




**Table 2: Statistics of the Four datasets.**

Domains	#Subjects	# Predicates	# Conflicting Predicates
Person	130174	16245	7506
Locations	74015	14162	6870
Organizations	25051	13956	6360
Descriptors	10362	6980	3250

We select three well-known state-of-the-art truth discovery methods as baseline including Vote, TruthFinder [8] and F-Quality Assessment [6]. The parameters of the baseline methods are set according to the authors’s suggestions. The experiments are performed on a desktop computer with Intel Core i5-3470 CPU 3.2 GHz with 4 GB main memory, and Microsoft Windows 7 professional operating system. All baseline methods were executed in the Eclipse (Java) platform by a single thread. Figure 4 shows the experimental results of all the methods in terms of accuracy in the four datasets.

**Figure 4: Performance comparison in the Four datasets**

It can be concluded from Figure 4 that TruthDiscover outperforms the three baseline methods in terms of accuracy. The main reason for this superiority is that it’s difficult for three baseline methods to estimate the reliability degree of “small” sources accurately in Linked Data, although these methods achieved satisfactory accuracy in certain datasets or applications.

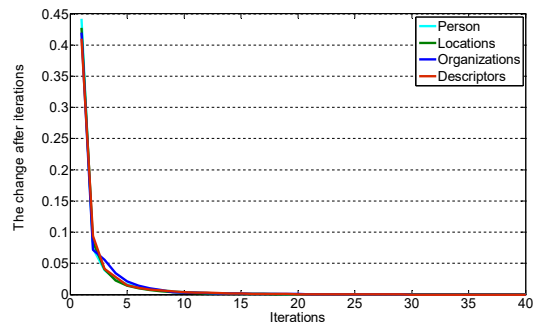
The experimental results of the average change in the trust value of objects after each iteration are shown in Figure 5.

It also shows the change decreases rapidly in the first five iterations, and then reaches a stable stage until the convergence criterion is satisfied.

## 5. CONCLUSIONS

We have developed an easy-to-use system called TruthDiscover<sup>8</sup>, which leverages the topological properties of the Source Belief Graph and the interdependencies between objects to infer the trustworthiness of sources and the trust values of

<sup>8</sup>Demonstration is available online at <http://123.139.159.38:9218/Truth/>

**Figure 5: Change in the trust values of objects after each iteration**

objects. This system is capable of automatically identifying the truth in massive Linked Data with a scale-free property. The experimental results show that TruthDiscover exhibits satisfactory accuracy. The future extension of TruthDiscover should include improving the performance of identifying the truth. This improvement can be achieved by identifying the copying relations of different sources.

## 6. ACKNOWLEDGEMENT

This work is funded by the National Key Research and Development Program of China (Grant No. 2016YFB1000903), the MOE Research Program for Online Education (Grant No. 2016YB166) and the National Science Foundation of China (Grant Nos. 61370019, 61672418, 61532004, 61532015).

## 7. REFERENCES

- [1] C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pages 205–227, 2009.
- [2] A. Dutta, C. Meilicke, and S. P. Ponzetto. A probabilistic approach for integrating heterogeneous knowledge sources. In *The Semantic Web: Trends and Challenges*, pages 286–301. Springer, 2014.
- [3] M. Fabian, K. Gjergji, and W. Gerhard. Yago: A core of semantic knowledge unifying wordnet and wikipedia. In *WWW, Alberta, Canada*, pages 697–706, 2007.
- [4] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han. Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. In *ACM SIGMOD*, pages 1187–1198, Utah, USA, 2014.
- [5] W. Liu, J. Liu, Y. Qian, B. Wei, and Q. Zheng. Truth discovery to resolve object conflicts in linked data. *arXiv preprint arXiv:1509.00104*, 2015.
- [6] J. Michelfeit, T. Knap, and M. Nečaský. Linked data integration with conflicts. *arXiv preprint arXiv:1410.7990*, 2014.
- [7] M. Schmachtenberg, C. Bizer, and H. Paulheim. State of the lod cloud 2014. *University of Mannheim, Data and Web Science Group*, 30, 2014.
- [8] X. Yin, J. Han, and P. S. Yu. Truth discovery with multiple conflicting information providers on the web. *IEEE Transactions on Knowledge and Data Engineering*, 20(6):796–808, 2008.