

In Search of Truth (on the Deep Web)

Divesh Srivastava

AT&T Labs Inc.
divesh@research.att.com

Abstract. The Deep Web has enabled the availability of a huge amount of useful information and people have come to rely on it to fulfill their information needs in a variety of domains. We present a recent study on the accuracy of data and the quality of Deep Web sources in two domains where quality is important to people's lives: Stock and Flight. We observe that, even in these domains, the quality of the data is less than ideal, with sources providing conflicting, out-of-date and incomplete data. Sources also copy, reformat and modify data from other sources, making it difficult to discover the truth. We describe techniques proposed in the literature to solve these problems, evaluate their strengths on our data, and identify directions for future work in this area.