

# Leave No Valuable Data Behind: The Crazy Ideas and the Business

Xin Luna Dong  
Amazon

lunadong@amazon.com

## ABSTRACT

With the mission "leave no valuable data behind", we developed techniques for knowledge fusion to guarantee the correctness of the knowledge. This talk starts with describing a few crazy ideas we have tested. The first, known as "Knowledge Vault", used 15 extractors to automatically extract knowledge from 1B+ Webpages, obtaining 3B+ distinct (subject, predicate, object) knowledge triples and predicting well-calibrated probabilities for extracted triples. The second, known as "Knowledge-Based Trust", estimated the trustworthiness of 119M webpages and 5.6M websites based on the correctness of their factual information. We then present how we bring the ideas to business in filling the gap between the knowledge at Google Knowledge Graph and the knowledge in the world.

## BIOGRAPHY

**Xin Luna Dong** is a Principal Scientist at Amazon, leading the efforts of constructing Amazon Product Graph. She was one of the major contributors to the Knowledge Vault project, and has led the Knowledge-based Trust project, which is called the "Google Truth Machine" by Washington's Post. She has co-authored book "Big Data Integration", published 65+ papers in top conferences and journals, given 20+ keynotes/invited-talks/tutorials, and got the Best Demo award in Sigmod 2005. She is the PC co-chair for Sigmod 2018 and WAIM 2015, and serves as an area chair for Sigmod 2017, Sigmod 2015, ICDE 2013, and CIKM 2011.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>. For any use beyond those covered by this license, obtain permission by emailing [info@vlldb.org](mailto:info@vlldb.org).

*Proceedings of the VLDB Endowment, Vol. 9, No. 13  
Copyright 2016 VLDB Endowment 2150-8097/16/09.*