Ontology Summit 2014 Session 11 Track D: Tackling the Variety Problem in Big Data – II

Ken Baclawski

Anne Thessen

Track D Co-Champions

Session Outline

- Ken Baclawski and Anne Thessen - Introduction

- Anatoly Levenchuk - Hackathon Announcement

- Mark Fox (University of Toronto) - **Variety in Big Data: A Cities Perspective**

- Malcolm Chisolm (AskGet.com) - **Data Governance to Manage Variety in Big Data**

- Dan Brickley (Google) - **Schema.org, FOAF and Linked Data: Lessons for Web-scale vocabulary deployment**

- Rosario Uceda-Sosa (IBM) - **Big Data, Open Data and the Smart City**

- Open Panel Discussion

Characteristics of Big Data

- Large amounts of data (volume)

- Rapid pace of data acquisition (velocity)

- Complexity of the data (variety)
    - Complexity of individual data items
    - Both structured and unstructured data
    - Complexity of data from one source
    - Multiple sources

What can ontologies do to help?

- Background knowledge of the domain

- The structure of the data

- Annotation of data and metadata

- Provenance of the data
    - Transformations
    - Analyses
    - Interpretations

- Data processing workflows

- Privacy concerns

- Hypothesis generation and their workflows

Still more opportunities for helping

- Data governance (Malcolm Chisholm)

- Data integration (Mark Fox and Rosario Uceda-Sosa)

- Schema.org, FOAF, Linked Data (Dan Brickley)

- ...

Speakers

- Mark Fox (University of Toronto)

  - More than half the world population live in cities and the proportion is growing, so cities are an enormous source of data. However, it is not just the amount of data that is daunting, but the enormous variety not only within a single city but also among the thousands of different cities.

- Malcolm Chisolm (AskGet.com)

  - Data governance is "a collection of disciplines that ensure data is managed adequately in an enterprise." Malcolm will discuss what is involved in data governance and its implications for managing Variety in Big Data.

Speakers

- Dan Brickley (Google)

  - Examines the Variety problem on the Web and provides an overview of some lessons learned from schema.org, FOAF and Linked Data deployment of RDF-based vocabularies.

- Rosario Uceda-Sosa (IBM)

  - Continues the examination of the Variety problem for city data and proposes some solutions.

Track D Wiki Pages

Synthesis Page

http://ontolog.cim3.net/cgi-bin/wiki.pl?
OntologySummit2014_Tackling_Variety_In_BigData_Synthesis

Community Input Page

http://ontolog.cim3.net/cgi-bin/wiki.pl?
OntologySummit2014_Tackling_Variety_In_BigData_CommunityInput