**BIG GEO-**DATA

*(from 30000 ft)*

Krzysztof Janowicz
University of California, Santa Barbara, USA
June 6, 2012

Thinking Spatially – Promoting spatial reasoning and technologies for research, creativity, and applications
IBM on Data: The New Natural Resource

Big Data

Big Data is the digital convergence of structured data found inside databases, and unstructured data flowing from new sources like social networks, mobile devices, sensors, RFID, smart meters and financial systems. Today, organizations can capture and analyze any data, regardless of what type, how much, or how fast it is moving, and make more informed decisions based on that information.

Big Data is growing fast

Annual growth rate

60%

Structured and unstructured data

NASA ON DATA: *Data-Model Intercomparison*

‘Considerable data regarding the environment are available through the myriad of remote-sensing programs, however, this data is not directly compatible with the models. It has been observed that scientists and engineers spend more than 60% of their time just preparing the data for model input or data-model intercomparison. This is an inefficient use of the precious time of NASA scientists and engineers.’ [NASA A.40]
**Big Geo-Data – Volume, Velocity, Variety**
**BIG GEO-DATA – VOLUME, VELOCITY, VARIETY**
Big Geo-Data – Volume, Velocity, Variety
ON-THE-FLY INTEGRATION AND CROSS-MEDIA EVALUATION

- Pictures, video, observations, simulations, news, reports from citizens & spokespersons,...
**Data Silos**

- Web services
- Databases
- Web pages
- hinder ad-hoc combination
- enforce data models
- limit re-usability
Use Uniform Resource Identifiers (URI) to identify entities, link them to other entities, encode information about these entities using the machine-understandable RDF, and make them available on the Web.
LINKED DATA – CREATING THE GLOBAL GRAPH
LINKED DATA – CREATING THE GLOBAL GRAPH
Searching the Web of Documents

Google search for "Populated places with population over 100000 near Mississippi River" shows about 20,600,000 results in 0.35 seconds.

- Illinois - Wikipedia, the free encyclopedia
  en.wikipedia.org/wiki/Illinois
- Largest cities in the United States by population by decade...
  en.wikipedia.org/.../Largest_cities_in_the_United_States_by_populati...
- Natural Earth
  www.naturalearthdata.com/

More search results available.
Motivation

Linking Geo-Data

Semantics

Interfaces

The Global Graph of Data

Searching the Web of Data

Populated places have a population, are located, occupy a certain area,...
Google’s Knowledge Graph

Population, Los Angeles, CA

Los Angeles - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Los_Angeles
Location within Los Angeles County in the state of California. Because of Los Angeles' large multi-ethnic population, a wide variety of faiths are practiced...

List of largest California cities by population - Wikipedia, the free...
en.wikipedia.org/wiki/List_of_largest_California_cities_by_population
List of largest California cities by population. From Wikipedia, the... Los Angeles is one of the world's centers of media, business, and international trade. It's the...

Los Angeles County, California - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Los_Angeles_Count...California
Los Angeles County (incorporated as the County of Los Angeles) is a county in the U.S. state of California. As of the 2010 census, the county had a population of...

Los Angeles County QuickFacts from the US Census Bureau
EXPLOITING THE GRAPH

About 2286 years (BC 276)
Eratosthenes, age
Hiko details
Eratosthenes - Biography of - about.com
Eratosthenes was born around 276 B.C.E. at a Greek ...
Eratosthenes - windows2universe.org
Eratosthenes was born in 276 B.C. ...
Is this accurate? Yes - No

Eratosthenes - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Eratosthenes
For the ancient Athenian statesman of the fifth century BC, see Eratosthenes .... 285 BC) since he is unlikely to have studied under him at the young age of 14.
Sieve of Eratosthenes - Eratosthenes (crater) - Eratosthenes Saamount

Eratosthenes - Biography of Eratosthenes
genography.about.com/cd/historyofgeography/a/eratosthenes.htm
Eratosthenes was born around 276 B.C.E. at a Greek colony in Cyrene, Libya. ... In old age, Eratosthenes became blind and died of self-induced starvation in ...

Eratosthenes of Cyrene was a Greek mathematician, geographer, poet, athlete, astronomer, and music theorist. He was the first person to use the word "geography" and invented the discipline of geography as we understand it.
Wikipedia
Born: BC 276, Cyrene
Died: BC 194, Alexandria
Parents: Agiaus

People also search for
Ptolemy
Aristarchos
Archimedes
Euclid
Hipparchos
EXPLOITING THE GRAPH

Different models, different inference → when do you stop counting?
Unfortunately, our data sources use exactly the same terminology (e.g., connection) to talk about totally different and contradicting facts (e.g., separation). While we can still syntactically integrate and reuse information, the results may be misleading or even meaningless.

A promising idea: when sharing data, define what you mean by terms such as city or river and relations such as population, or date of birth to avoid misunderstandings and foster interoperability.
Unfortunately, our data sources use exactly the same terminology (e.g., connection) to talk about totally different and contradicting facts (e.g., separation). While we can still syntactically integrate and reuse information, the results may be misleading or even meaningless.

**A promising idea:** when sharing data, define what you mean by terms such as city or river and relations such as population, or date of birth to avoid misunderstandings and foster interoperability.
**Semantic Annotation and Ontologies**

- **Ontologies** restrict the interpretation of domain vocabulary towards their intended meaning
DiaLoc: New Data Sources for Old Questions

- 'market food street narrow dense populated asia economy air conditioning smog fog humid warm building construction skyscrapers skyline shipping export channel harbor transportation tram city advertisement'

- **Variety**: Plain text, not image features as data source
Automatically extract relevant features while keeping others invariant.