Use Case - The Web Observatory

(Contributed by Adriano C. Machado Pereira, Adriano Veloso, Gisele Pappa, Wagner Meira Jr.)
URL: http://observatorio.inweb.org.br/english.html

There are almost 65 million brazilians connected to the Internet - 36% of the Brazilian Population, according to Comitê Gestor da Internet no Brasil. As a consequence, events such as the Brazilian Election Running have become popular topics in the Web, mainly in Online Social Networks. Our goal is to understand this new reality and present new ways to watch facts, events and entities on the fly using the Web and user-generated content available in Online Social Networks and Blogs.

The **Web Observatory** is a research project part of the Instituto Nacional de Ciência e Tecnologia para a Web (INWEB), sponsored by CNPq and Fapemig. There are over 30 experts involved in the project, from four differente Federal Universities: Universidade Federal de Minas Gerais (UFMG), Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Universidade Federal do Amazonas (UFAM) e Universidade Federal do Rio Grande do Sul (UFRGS).

The INWEB researchers use a set of new techniques related to Information Recovery, Data Mining and Data Visualization to understand and summarize what the media and users are talking about on the Web. That is the fundamental piece to evaluate the impact of the Olympic Campaigns and how users react to news and discussions. One new feature in this project is the possibility to see the propagation of the Tweets

Elements:

- ***Domains:*** Different contexts or domains, related to data from the Web. For example:
  - *Health (for example, diseases);*
  - *Tourism;*
  - *Sports (for example, soccer championship and Olympic games)*
  - *Politics;*
  - *Finance;*
  - *Etc.*
- ***Obligation/motivation:*** Data must be obtained from different public data sources from the Web.
- ***Usage:*** Provide different data analysis, indicators or visualizations to allow a better understand of a context.
- ***Quality:*** Variable, depend on the data source, can be structured or not.
- ***Size:*** Variable, can be small data instances to a huge amount of data, depending on the context under investigation. In general, there are a huge amount of data.
- ***Type/format:*** Diverse, like CSV, HTML, JSON, XML, etc.
- ***Rate of change:*** different rates of change, usually very dynamic.
- ***Potential audience:*** Diverse, different Web users.

Challenges:

- Data volume;
- Data velocity;
- Data variety;
- Data value;
- Complexity;

Requires: R-MetadataMachineRead , R-MetadataStandardized , R-MetadataDocum , R-VocabReference , R-VocabDocum , R-VocabOpen , R-ProvAvailable, R-GranularityLevels

Requires(new): R-DataCharacterization, R-DataImputation, R-DataSegmentation, R-DataDisambiguation, R-DataEntityRecognition, R-DataFusion

--

Challenge	Requirements	
Data Enrichment	Requirements for Data Enrichment	R-DataCharacterization, R-DataImputation, R-DataSegmentation, R-DataDisambiguation, R-DataEntityRecognition, R-DataFusion