Big data serving towns and villages

We live in an information age. According to IBM, 90% of all data in the world has been generated over the last two years*, providing us with a vast knowledge base. How do we obtain meaningful information from this enormous pool? How do we sort, analyze and interpret it so we can benefit from it in our daily life?

In this connection, ways of using and visualizing data were the central topic of a lecture recently presented at the Technical University of Košice and Masaryk University in Brno, by Šimon Urbánek, from AT&T Labs.

Every day, a huge amount of data is generated in our world. The power industry, telecommunications, and finance-sector corporations, as well as public administration authorities and government organizations, generate and collect copious amounts of data. Ranging in type, structure and size, the data is incessantly generated in large amounts, which is why it has been termed "Big Data".

Some of the characteristics of Big Data include size, rapid growth, and diversity. At the same time, such data presents opportunities for us to enhance our understanding of human behavior or make predictions regarding the future development in certain spheres of life. Knowledge obtained in this way can be utilized both commercially and non-commercially, e.g. in the sphere of services, such as public sector services or – as pointed out by Šimon Urbánek – those related to transport planning or public space usage in cities. Analyzing, processing and visualizing big data can be carried out by means of open-source tools, such as *Nanocubes* or *RCloud*, which are instrumental in processing and evaluating large datasets.

Nanocubes offers interactive visualizations of billions of datasets in real time. Users can choose various criteria for data sorting, such as space and time, and display their data via a web browser using heat maps, bar graphs or histograms.

To illustrate the functionality of the software, the www.nanocubes.net website offers an interesting, practical example providing a visualization of crime data related to the city of Chicago. For example, users have a chance to see whether seasonal crimes are prevalent, which areas are affected by selected crimes, or how crime rates develop in time.

> **90% of all data in the world has been generated over the last two years.**

Šimon Urbánek also presented a visualization of tweets that is based on 35 Terabytes of source data obtained over a period of nine months, allowing the user to apply filters involving geographical coordinates, device types, visited pages, or different languages. These types of data provide very useful guidelines for internet connection providers because they can easily anticipate increased network usage, identify days requiring higher capacities, or determine the degree to which higher capacities are needed.

The usefulness of data visualization for urban planners is exemplified by the case of Morristown, New Jersey. Based on data obtained from transmitters, such as the number of phone calls and text messages, the most frequent routes people used to get into town could be easily mapped. The results of the research enabled urban planners to make adjustments to the transport network of the city.

Available at github.com/att/rcloud, *RCloud* represents a platform for analyzing bulky datasets based on the programming language called "R". Similarly to *Nanocubes*, the website illustrates the potential of large dataset analysis using practical examples.

Given the recent development, it can be assumed that the amount of big data will keep on growing. As a result, commercial companies, public institutions and organizations are likely to be increasingly motivated to make the best of such data either using *Nanocubes* or *RCloud* – or any other suitable tools for that matter.

*Filip Šváb
Executive Director,
International External
Affairs, AT&T*

* IBM.com – Apply news analytics tools to reveal new opportunities: http://www.ibm.com/smarterplanet/us/en/business_analytics/article/it_business_intelligence.html

Šimon Urbánek from AT&T Labs talking to students at the Technical University of Košice