MTA SZTAKI

Informatics Laboratory

Head of department:
Dr. Andras Benczur

Phone:
+36 1 279 6172

E-mail:
andras. benczur@sztaki.mta.hu

Address:
H-1111 Budapest, Lágymányosi u. 11.

Web:
infolab.sztaki.mta.hu/en

INTRODUCTION

"Big Data" is an emerging new research area for the methodologies of extreme large scale problems in business intelligence, e-science and Web mining. We concentrate on applications for social network mining, graph clustering, personalized and similarity search, recommendation and spam filtering, as well as security problems ranging from financial risk analysis or insurance fraud to people trafficking or organized crime.

MAIN R&D TOPICS

Society has reached a point of no return, one that leaves us completely reliant on omnipresent ICT-mediated communication. Mobile and sensor-rich portable devices connect millions of humans with Petabytes of data and numerous on-line services. However, tearing down the physical-digital barrier in a scalable fashion requires both radically novel algorithmic knowledge and in-depth understanding of humans and societies. We will deliver major theoretical advances in real-time intelligent information management of large datasets including online social networks, mobile devices and humans in physical space by delivering three functions: "alert", by real-time location-aware knowledge acquisition, analysis and visualization; "response", through on-demand composition and coordination of large teams; and effective "communication", through recommendation and personalization.

We conduct research ranging from theory to experimentation by building on the unique nature of our research lab. We cover the full chain from core research to industrial deployment, including unique access to data ranging from telecommunication logs to large scale Web crawls. As a particular strength in our previous results, we design algorithms that handle the explosive growth in data sizes and impose no artificial size limits for real-world applications. The highlights of our proposed research with both novel areas as well as related fields where we have the strongest existing results that are listed next.

The R&D results of the laboratory focus on data mining and search solutions for community and link analysis, custom solutions for extreme large systems (large Intranets, high traffic portals) as well as for languages with particularly complex syntax in collaboration with computational linguistic groups. For the quality of our research results we were awarded a *Yahoo! Faculty Research Grant* in the academic year 2006/2007. In 2007 the Group achieved First Prize on the prestigious KDD Cup, a competition involving the best data mining groups around the world. Several of our former PhD students work now at the research centres of the leading internet search companies.

International scientific partnerships

Portugal Telecom, NMusic (PT), Rovio (FI, the maker of the Angry Birds games) and Internet Memory Research Paris as industrial research centres in collaboration with DFKI, Max-Planck Institut für Informatik, SICS, CNRS and among others the TU Berlin and Universities of Milan, Twente and Patras.

Our know-how includes machine learning applied in a wide range of areas of industry, finance, telecommunications, web and social media, as well as core system and data analytics technologies over distributed batch and data streaming systems, including Flink, Spark and Hadoop, to name a few.

- Yahoo! Faculty Research Grant, 2006
- KDD Cup 2007, 2009
- VAST Challenges 2009-2012
- ImageCLEF 2011, 2012
- RecSys Challenges 2014, 2015

- Ericsson Hungary: Data analytics and distributed Big Data stream processing, 2014-.
- OTP Bank: Machine learning tools for modelling in finance, 2015-.
- Bosch Hungary: Machine learning for production quality prediction, 2015-.
- AEGON Hungary: Data warehousing, customer data deduplication and network analysis, fraud detection, 2006-.

- In the STREAMLINE H2020: Improving Competitiveness of European Enterprises through Streamlined Analysis of Data at Rest and Data in Motion project we build appropriate data stream-oriented analytics tools that reduce the complexity, cost, and burden associated with jointly supporting analytics for both "data at rest" and "data in motion."

- In the NADINE FP7 288956: New tools and Algorithms for DIrected NEtwork analysis project we have developed new tools and algorithms that will create a fundamental basis for developers of new types of search and social media services.