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 online 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.

“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.

Research and development activities

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 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 centers of the leading internet search companies.
National References
- Web server and IT log analytics system, T-Online, AEGON Hungary.
- Hungarian Telecom: Call community and customer analysis tool.
- AEGON Hungary: Desktop search engine, data warehousing, customer network analysis, fraud detection, car insurance campaign toolkit, 2006-.

International References
- Yahoo! Faculty Research Grant, 2006
- KDD Cup 2007, 2009
- VAST Challenges 2009-2012
- ImageCLEF 2011, 2012

International research partners
Yahoo Research Barcelona and Internet Memory Research Paris as industrial research centers in collaboration with Max-Planck Institut für Informatik, CNRS and among others the Universities of Milan, Twente and Patras.

Know-how and industrial solutions
Major software products include a customer relation management software capable of visualizing the connection between entities (persons, objects, contracts) as well as in a search engine with integrated linguistic tools for the Hungarian language that serves the Intranet of national branches of multinational companies (T-Mobile, Vodafone, AEGON). We led several customer relation management and risk assessment projects for AEGON Hungary.

Participation in EU projects
- In the LAWA FP7 258105: Longitudinal Analytics of Web Archive data project we build an Internet-based experimental testbed for large-scale data analytics.
- The SCIIMS FP7 218223: Strategic Crime and Immigration Information Management System project aids combating organized crime with focus on people trafficking and smuggling.
- In the NADINE FP7 288956: New tools and Algorithms for Directed NETwork analysis project we develop new tools and algorithms produced by the project will create fundamental basis for developers of new types of search and social media services, which will put Europe on leading positions.

Education
Participation in masters and doctoral education at the Technical University and Eötvös University, Budapest in algorithms, data mining and networks.

People
András Benczúr, head
Lajos Rónyai, János Demetrovics, members of the HAS
Dániel Marx, ERC Starting Grant winner
Over 30 researchers, including 11 PhD-s, 8 doctoral students and 7 undergraduates.