



## Informatikai Kutatólaboratórium

Vezető  
**Dr. Benczúr András**

Telefon:  
**+36 1 279 6172**

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

Cím:  
**H-1111 Budapest, Lágymányosi u. 11.**

Web:  
**infolab.sztaki.mta.hu**

### BEMUTATKOZÁS

A „Big Data” feladatokban a problémát az adat mérete maga jelenti. Az üzleti intelligencia, a Web és szenzorhálózatok adatbányászata területén jelentkező extrém méretű információfeldolgozási problémák között kiemelten vizsgáljuk a szociális közösségeket és jelenségeket, a személyre szabott információ keresését, ajánló rendszereket, spam szűrőket és a navigációs, mobilitási adatok időbeli dinamikáját.

### FŐBB K+F IRÁNYOK

Társadalmunk számára már nincs visszalépés: az életünk minden területét átszövik az infokommunikációs technológiák. Szenzorok sokaságával felszerelt mobil eszközök felhasználók százmillióit kötik össze és adatok Petabyte-jait termelik naponta. A fizikai és digitális világ közötti választóvonal áttöréséhez azonban radikálisan új, skálázható algoritmikus technikák és az egyének és a társadalom működésének ismerete szükséges.

A laboratórium célja, hogy zárt láncban teljes humán és hagyományos innovációs folyamatot valósítson meg a képzéstől (ELTE, BME), doktori munka vezetésétől a publikációs tevékenységen keresztül az eredmények ipari hasznosításáig. Talán országosan is egyedülálló a matematikusi és mérnöki munka együttélése: a kutatás alapvetően kísérleti jellegű, a módszerek létjogosultságát az igazolja, ha valós adatokon és alkalmazásokban hasznosítható eredményt, információt, új tudást hoznak létre. Ugyanakkor az adatok extrém mérete miatt minden lehetséges eljárás mély algoritmuselméleti és valószínűségi számítási ismereteken, matematikailag bizonyítható alapokon kell, hogy álljon.

A laboratórium K+F eredményei egyedi megoldásokat biztosítanak extrém méretű rendszerek (nagy belső hálózatok, nagyforgalmú Web portálok), illetve a magyar nyelvű információkeresés számára. A csoport hazai és nemzetközi együttműködésben szociális hálózatok, biztonsági incidensek és ügyfeladatok elemzését, illetve a Web spam szűrését és ajánló rendszerek fejlesztését végzi; eredményeit a 2006/07 akadémiai évben Yahoo! Faculty Research Grant támogatással ismerték el, illetve a KDD Cup 2007 győztes megoldása is a csoport nevéhez fűződik. A végzett doktoranduszok közül már öten dolgoznak a legnagyobb keresőcégek kutató-fejlesztő központjaiban.

Gépi tanulási megoldásokat biztosítunk széles alkalmazási területen, beleértve a gyártás, pénzügy, telekommunikáció, web és közösségi média adatelemzési feladatait. Partnereink számára olyan elosztott adatelemző platformokon, mint például a Flink, Spark vagy Hadoop, alap infrastruktúra optimalizálás és gépi tanulási módszerek kutatás-fejlesztését végezzük.

- Yahoo! Faculty Research Grant, 2006
- KDD Cup 2007, 2009
- VAST Challenges 2009-2012
- ImageCLEF 2011, 2012
- RecSys Challenges 2014, 2015
- Ericsson Magyarország: Távközlési adatok prediktív elemzése és elosztott Big Data adatfolyam feldolgozás
- OTP Bank: gépi tanulás a pénzügyi elemzés területén
- Bosch Hungary: gépi tanulás a gyártás-minőség előrejelzésre
- AEGON: Adattárház és ügyfél deduplikáció, Biztosítási csalás és biztosítás-árrugalmasság elemzés
- A STREAMLINE H2020: Improving Competitiveness of European Enterprises through Streamlined Analysis of Data at Rest and Data in Motion projekt célja az európai alternatíva és általában az európai Data Science megteremtése a DFKI által irányított Apache Flink platformra alapozva. Feladatunk a Data Streaming komponens tervezése, kialakítása, és hatékony algoritmusok tervezése, különös tekintettel az Ericsson együttműködésben végzett mobil szenzor adatokat elemző és a valós idejű ajánló alkalmazásokra.
- A NADINE FP7 288956: New tools and Algorithms for DIrected NEtwork analysis projekt célja új kereső és közösségi szolgáltatások alapjainak megteremtése, amelyek az Európában működő tartalomszolgáltató ipar versenyképességét fokozzák.

