Véleményformálók és érzelmi analízis közösségi hálózatokon

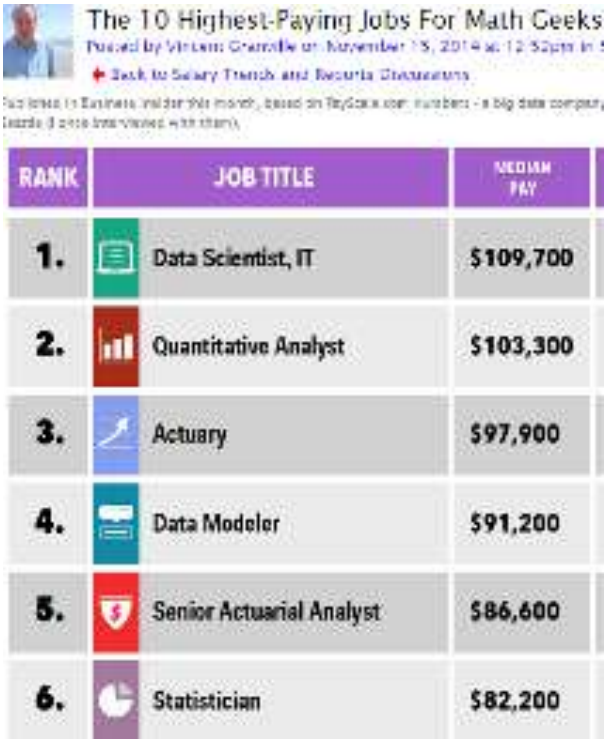
témavezető: **Kiss Attila**
kiss@inf.elte.hu
ELTE IK, Információs Rendszerek Tanszék, D2.508

Nagy adathalmazok (Big Data) elemzését összefoglalóan adattudománynak (Data Science) nevezik.

Ebbe a tudományágba beletartozik az adatok gyűjtése, tisztítása, tárolása, elemző algoritmusok készítése, adatbányászati technológiák felhasználása, statisztikai vizsgálatok, osztályozás, klaszterezés, gyakori minták keresése, gráfok, hálózatok elemzése.

A hatékony megoldásokra az adatközpontokban (Data Center) általában nagy valós, vagy virtuális gépparkon futó osztott rendszereket használnak.

Az adattudós állás 2016-ban az USA-ban több felmérés alapján (állásigény, kereset, juttatások, előrehaladási lehetőség) a legjobb állás.



The 10 Highest-Paying Jobs For Math Geeks

RANK	JOB TITLE	MEDIAN PAY
1.	Data Scientist, IT	$109,700
2.	Quantitative Analyst	$103,300
3.	Actuary	$97,900
4.	Data Modeler	$91,200
5.	Senior Actuarial Analyst	$86,600
6.	Statistician	$82,200

A témára két matematikus, alkalmazott matematikus vagy informatikus jelentkezését várom. (Statisztikai, aktuáriusi elméleti és gyakorlati ismeretek előnyt jelentenek.)

Sikeres előrehaladás, nyertes TDK dolgozat esetén, ipari környezetben versenyképes kiegészítő ösztöndíj is megpályázható.

Kiindulási feladat: Érzelmi analízis és véleményterjedés közösségi hálókon. Nyelvi elemzés után, megfelelően betanított modellt készítve, határozzuk meg, hogy például két termék, két politikus, két film közül melyik tetszik jobban az embereknek a fórumokon, blogokon, közösségi hálózatokon megtalálható információk alapján.

Második feladat: Kik a véleményformálók, akiknek a véleményét, ízlését a legtöbben átveszik a közösségi hálókon. Hogyan terjed az információ, a pletyka a hálózatokon?

További kutatási feladatok: a fenti területek bármelyikéhez (adatbázisok, adatbányászat, adattudomány, osztott feldolgozások) tartozó új, hatékony megoldások formalizálása, elemzése matematikai vagy informatikai szempontból, algoritmusok tervezése, implementálása, összehasonlítása már ismert megoldásokkal.

Kiindulási irodalom: a Stanford egyetem letölthető tankönyve: Mining of Massive Datasets

http://www.kdnuggets.com/2014/02/book-mining-massive-datasets-2nd-edition-free-download.html