

## A TANTÁRGY ADATLAPJA

### 1. A képzési program adatai

1.1 Felsőoktatási intézmény	<b>Babeş–Bolyai Tudományegyetem</b>
1.2 Kar	<b>Matematika és Informatika Kar</b>
1.3 Intézet	<b>Magyar Matematika és Informatika Intézet</b>
1.4 Szakterület	<b>Informatika</b>
1.5 Képzési szint	<b>Mesteri</b>
1.6 Szak / Képesítés	<b>Adatelemzés és modellezés</b>

### 2. A tantárgy adatai

2.1 A tantárgy neve	<b>Nagy adathalmazok elemzése</b> <b>Analiza datelor masive</b> <b>Analysis of massive datasets</b>						
2.2 Az előadásért felelős tanár neve	<b>Darvay Zsolt</b>						
2.3 A szemináriumért felelős tanár neve	<b>Darvay Zsolt</b>						
2.4 Tanulmányi év	<b>2</b>	2.5 Félév	<b>1</b>	2.6. Értékelés módja	<b>vizsga</b>	2.7 Tantárgy típusa	<b>kötelező – alap</b>
2.8 A tantárgy kódja	<b>MMM8075</b>						

### 3. Teljes becsült idő (az oktatási tevékenység féléves óraszama)

3.1 Heti óraszám	<b>5</b>	melyből: 3.2 előadás	<b>2</b>	3.3 szeminárium/labor/praktika	<b>3</b>
3.4 Tantervben szereplő össz-óraszám	<b>70</b>	melyből: 3.5 előadás	<b>28</b>	3.6 szeminárium/labor	<b>42</b>
A tanulmányi idő elosztása:					Óra
A tankönyv, a jegyzet, a szakirodalom vagy saját jegyzetek tanulmányozása					<b>49</b>
Könyvtárban, elektronikus adatbázisokban vagy terepen való további tájékozódás					<b>14</b>
Szemináriumok / laborok, házi feladatok, portfóliók, referátumok, esszék kidolgozása					<b>49</b>
Egyéni készségfejlesztés (tutorálás)					<b>14</b>
Vizsgák					<b>4</b>
Más tevékenységek:					
3.7 Egyéni munka össz-óraszama					<b>130</b>
3.8 A félév össz-óraszama					<b>200</b>
3.9 Kreditszám					<b>8</b>

### 4. Előfeltételek (ha vannak)

4.1 Tantervi	Nincs
4.2 Kompetenciabeli	Alapvető algoritmusok, programozási készségek, matematikai alapismeretek (algebra, valószínűségszámítás).

### 5. Feltételek (ha vannak)

5.1 Az előadás lebonyolításának feltételei	<ul style="list-style-type: none"> <li>Az előadásokhoz videoprojektor szükséges.</li> <li>A példák kifejtéséhez és illusztráció számára tábla szükséges.</li> </ul>
5.2 A szeminárium / labor lebonyolításának feltételei	<ul style="list-style-type: none"> <li>A laboratóriumi órák alatt a diákok a számítógépet, az oktató a táblát használja.</li> </ul>

## 6. Elsajátítandó jellemző kompetenciák

<b>Szakmai kompetenciák</b>	<ul style="list-style-type: none"> <li>• Nagy adathalmazok elemzése információkinyerés céljából</li> <li>• Adatbányászati algoritmusok elemzése és fejlesztése</li> </ul>
<b>Transzverzális kompetenciák</b>	<ul style="list-style-type: none"> <li>• Önálló tanulás</li> <li>• Munkamódszerek, módszertani kompetenciák</li> <li>• Kritikus gondolkodás és reflexió</li> </ul>

## 7. A tantárgy célkitűzései (az elsajátítandó jellemző kompetenciák alapján)

7.1 A tantárgy általános célkitűzése	<ul style="list-style-type: none"> <li>• A tantárgy célja a nagy adathalmazok elemzéséhez, feldolgozásához szükséges módszerek bemutatása.</li> </ul>
7.2 A tantárgy sajátos célkitűzései	<ul style="list-style-type: none"> <li>• Az elemzési módszerek fogalmainak és algoritmusainak ismerete: <ul style="list-style-type: none"> <li>○ Adatbányászati alapfogalmak</li> <li>○ Társítási szabályok bányászata</li> <li>○ Legközelebbi szomszédok gyors keresése</li> <li>○ A MapReduce modell</li> <li>○ Link-analízis</li> <li>○ Tanuló algoritmusok nagy adathalmazokon</li> <li>○ Adatbányászat adatfolyamokból és gráfokból</li> </ul> </li> </ul>

## 8. A tantárgy tartalma

8.1 Előadás	Didaktikai módszerek	Megjegyzések
<b>1.</b> Adatbányászati fogalmak, statisztikai modellezés, a Bonferroni-elv.	tanári magyarázat, munkáltatás	
<b>2.</b> Társítási szabályok bányászata nagy adatbázisokban: vásárlói kosár elemzése, az Apriori algoritmus.	tanári magyarázat, munkáltatás	
<b>3.</b> Legközelebbi szomszédok hatékony keresése: a Locality-Sensitive Hashing (LSH) metódus.	tanári magyarázat, munkáltatás	
<b>4.</b> Az LSH módszer elmélete.	tanári magyarázat, munkáltatás	
<b>5.</b> Osztott fájlrendszerek és a MapReduce modell.	tanári magyarázat, munkáltatás	
<b>6.</b> MapReduce algoritmusok és a MapReduce kiterjesztései.	tanári magyarázat, munkáltatás	
<b>7.</b> Tanuló algoritmusok nagy adathalmazokon: k-legközelebbi szomszéd módszere (kNN), perceptron algoritmus.	tanári magyarázat, munkáltatás	
<b>8.</b> Tanuló algoritmusok nagy adathalmazokon: szupport vektor gépek.	tanári magyarázat, munkáltatás	
<b>9.</b> Link-analízis: PageRank, topic-sensitive PageRank, link spam, HITS (Hubs and Authorities).	tanári magyarázat, munkáltatás	
<b>10.</b> Adatbányászat adatfolyamokból.	tanári magyarázat,	

	munkáltatás	
<b>11. Klaszterezés elemzések.</b>	tanári magyarázat, munkáltatás	
<b>12. Dimenzióredukciós módszerek: az SVD és CUR mátrixdekompozíciók.</b>	tanári magyarázat, munkáltatás	
<b>13-14. Adatbányászat gráfokból: szociális háló elemzése, szociális háló klaszterezése, közösségek felfedezése, gráfparticionálás, a Simrank algoritmus, háromszögek számlálása, gráfok szomszédsági tulajdonságai.</b>	tanári magyarázat, munkáltatás	

### **Könyvészet**

- [1] RAJARAMAN A., LESKOVEC J., ULLMAN J.D. *Mining of Massive Datasets*. Cambridge University Press, 2011.
- [2] BOTTOU L., CHAPPELLE O., DECOSTE D., WESTON J. *Large-Scale Kernel Machines*. MIT Press, 2007.
- [3] LIN J., DYER C. *Data-Intensive Text Processing with MapReduce*. Morgan & Claypool Publishers, 2010.
- [4] BALDI P., FRASCONI P., SMYTH P. *Modeling the Internet and the Web. Probabilistic Methods and Algorithms*. Wiley, 2003.
- [5] COOK D.J., HOLDER L.B. *Mining Graph Data*. Wiley, 2007.
- [6] SHAKHAROVICH G., DARRELL T., INDYK P. *Nearest-Neighbor Methods in Learning and Vision. Theory and Practice*. MIT Press, 2006.

<b>8.2 Szeminárium / Labor</b>	<b>Didaktikai módszerek</b>	<b>Megjegyzések</b>
<b>1. A véletlen hipersík alapú Locality-Sensitive Hashing módszer.</b>	munkáltatás, demonstráció, példák	
<b>2. Adatok indexelése.</b>	munkáltatás, demonstráció, példák	
<b>3. MapReduce alapú rendszerek: Hadoop.</b>	munkáltatás, demonstráció, példák	
<b>4. A PageRank algoritmus.</b>	munkáltatás, demonstráció, példák	
<b>5. Az SVD és CUR mátrixdekompozíciók.</b>	munkáltatás, demonstráció, példák	
<b>6. Az SVM optimalizálási feladatának megoldása ritka adatok esetén.</b>	munkáltatás, demonstráció, példák	
<b>7. Az SMO algoritmus.</b>	munkáltatás, demonstráció, példák	

### **Könyvészet**

- [1]–[6] +
- [7] LANGVILLE A.N., MEYER C.D. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2006.
- [8] PERERA S., GUNARATHNE T. *Hadoop MapReduce Cookbook*. Packt Publishing, 2013.
- [9] HAN J., KAMBER M. *Adatbányászat. Konceptiók és technikák*. Panem, 2004.
- [10] <http://www.stanford.edu/class/cs246/handouts.html>

**9. Az episztemikus közösségek képviselői, a szakmai egyesületek és a szakterület reprezentatív munkáltatói elvárásainak összhangba hozása a tantárgy tartalmával.**

- Az előadás felépítése megegyezik a Stanford-on oktatott "Mining massive data sets" c. tantárgyével (<http://www.stanford.edu/class/cs246/>).
- A kurzus fontos fogalmakat es algoritmusokat mutat be, melyek szükségesek a rohamos mértékben bővülő, nagy adathalmazok feldolgozásához.

**10. Értékelés**

<b>Tevékenység típusa</b>	<b>10.1 Értékelési kritériumok</b>	<b>10.2 Értékelési módszerek</b>	<b>10.3 Aránya a végső jegyben</b>
10.4 Előadás	Írásbeli vizsga a félév végén	Írásbeli vizsga	60%
10.5 Labor	Programozási feladatok bemutatása	A megoldások pontozása	40%
<b>10.6 A teljesítmény minimumkövetelményei</b>			
Kötelező a pontok felének összeszedése minden kiértékeléskor (évközi kiértékelés (laborgyakorlatok), végső vizsga).			

Kitöltés dátuma

Előadás felelőse

Labor / praktika felelőse

2018.04.22

Dr. Darvay Zsolt, adjunktus

Dr. Darvay Zsolt, adjunktus

Az intézeti jóváhagyás dátuma

Intézetigazgató

.....

Dr. András Szilárd, egyet. docens