

A TANTÁRGY ADATLAPJA

1. A képzési program adatai

1.1 Felsőoktatási intézmény	Babeş–Bolyai Tudományegyetem
1.2 Kar	Politika-, Közigazgatás- és Kommunikációtudományi Kar
1.3 Intézet	Újságírás és Digitális Média Intézet
1.4 Szakterület	Kommunikációtudományok
1.5 Képzési szint	BA
1.6 Szak / Képesítés	Digitális média

2. A tantárgy adatai

2.1 A tantárgy neve	Digitális adatelemzés						
2.2 Az előadásért felelős tanár neve	LECT. DR. Péter Árpád						
2.3 A szemináriumért felelős tanár neve	LECT. DR. Péter Árpád						
2.4 Tanulmányi év	II.	2.5 Félév	II.	2.6. Értékelés módja	vizsga	2.7 Tantárgy típusa	kötelező

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

3.1 Heti óraszám	3	melyből: 3.2 előadás	2	3.3 szeminárium/labor	1
3.4 Tantervben szereplő össz-óraszám	42	melyből: 3.5 előadás	28	3.6 szeminárium/labor	14
A tanulmányi idő elosztása:					óra
A tankönyv, a jegyzet, a szakirodalom vagy saját jegyzetek tanulmányozása					24
Könyvtárban, elektronikus adatbázisokban vagy terepen való további tájékozódás					24
Szemináriumok / laborok, házi feladatok, portofóliók, referátumok, esszék kidolgozása					24
Egyéni készségfejlesztés (tutorálás)					4
Vizsgák					4
Más tevékenységek: szimulációs gyakorlatok					3
3.7 Egyéni munka össz-óraszama	83				
3.8 A félév össz-óraszama	125				
3.9 Kreditszám	5				

4. Előfeltételek (ha vannak)

4.1 Tantervi	•
4.2 Kompetenciabeli	•

5. Feltételek (ha vannak)

5.1 Az előadás lebonyolításának feltételei	<ul style="list-style-type: none"> Videoprojektor+tanári számítógép a teremben
5.2 A szeminárium / labor lebonyolításának feltételei	<ul style="list-style-type: none"> Számítógépek a teremben Szükséges szoftverek: Helium Scaper, Chrome Web Scaper, Import io, MS Excel, Notepad++, NodeXL Excel Add-in, Facepager, Open Refine, Excel ASAP Utilities, Google Chrome, Mozilla Firefox, Tableau Public, MS Excel, MS Access, KH Coder

6. Elsajátítandó jellemző kompetenciák

Szakmai kompetenciák	<p>C2.1 Az információkezelés, a médiakommunikáció, a médiarendszerek, a médiadinamika és a médiakommunikáció hatásainak elméleteinek azonosítása és megértése.</p> <p>C2.3 Megfelelő technológiák használata a különböző termékek professzionális előállítása és kommunikációja során</p> <p>információs termékek, tiszteletben tartva a nyilvánosság pontos tájékoztatáshoz való jogát</p> <ul style="list-style-type: none"> - Képesség olyan szoftverek és szolgáltatások használatára, amelyek lehetővé teszik a webkaparók/hálótörlők felállítását; - Képes olyan programok és szolgáltatások használatára, amelyek lehetővé teszik a webszolgáltatások API-inak lekérdezését; - Képes nagy adathalmazok tisztítására és rendszerezésére szolgáló programok használatára <p>C2.4 Szabványos kritériumok és módszerek alkalmazása a különböző technológiák használatának minőségének és korlátainak értékelésére a konkrét információs és kommunikációs cselekményekben.</p> <p>C2.5 Portfólió készítése a szakmai információs és/vagy kommunikációs tevékenység médiadimenziójáról</p> <ul style="list-style-type: none"> - Képesség nagy adathalmazok felhasználására és összesítésére vizualizációk készítéséhez; - Képesség a webes jelenségek elemzésére az adatkészletek felhasználásával létrehozott vizualizációk alapján.
Transzverzális kompetenciák	<ul style="list-style-type: none"> - (CT1) Elméleti és gyakorlati módszerekkel hatékony alkalmazásával történő etikus megoldása a jellemző szakmai problémáknak.

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

7.1 A tantárgy általános célkitűzése	<ul style="list-style-type: none"> • - Az online környezetből származó nagy adathalmazok gyűjtésének és felhasználásának képességének fejlesztése
7.2 A tantárgy sajátos célkitűzései	<ul style="list-style-type: none"> • A webhez, a statikus és dinamikus weboldalakhoz, valamint a weboldalak működéséhez kapcsolódó alapfogalmak megértése; • Megérti, hogyan működnek a webes scraperek/ wrapperek és az API-k; • Az adatok főbb típusainak és feldolgozásuk módjának megértése; • Az adatvizualizációk készítésének és nagy adathalmazokhoz való igazításának alapelveinek megértése; • Fejlessze ki a nyílt forráskódú alkalmazások segítségével egy scraper/wrapper felállításának képességét; • Nagyméretű adathalmazok rendszerezésének és tisztításának képessége; • Nagy adathalmazok vizuális megjelenítésének képességének fejlesztése.

8. A tantárgy tartalma

8.1 Előadás	Didaktikai módszerek	Megjegyzések
1. A web mint adatforrás. Bevezetés a Web Wrapping/Scraping/Crawling - történelmébe és definícióiba	Frontális oktatás, multimédiás bemutató, beszélgetés	
2. Nyilvános adatok és nyílt adatok. Web Wrapping/Scraping/Crawling - jogi kérdések.	<i>idem</i>	
3. Az automatizált adatgyűjtés lépései webes adatgyűjtéssel. HTML sablonok. Navigáció szimulálása	<i>idem</i>	
4. Automatikus műveletek beállítása. Oldalszámzás. Hitelesítés. Többszintű kivonás	<i>idem</i>	
5. Adattípusok és adatszerkezetek. Lehetséges műveletek adattípusokkal. Adatok megjelenítése. Adatkonverzió.	<i>idem</i>	
6. Relációs adatbázisok - bevezető fogalmak. Adattípusok.	<i>idem</i>	
7. Adatbázisok. Táblák. Kapcsolatok. Lekérdezések SQL-ben - alapok.	<i>idem</i>	
8. BA közösségi média/generikus szolgáltatások API-inak lekérdezése	<i>idem</i>	
9. A webből kinyert adathalmazok tisztításának és kezelésének elvei és technikai - karakterláncműveletek, automatikus kategorizálás	<i>idem</i>	
10. Kulcsszóelemzés. Speciális változók, mérőszámok és elemzések a keresőmotor-optimalizáláshoz (SEO)	<i>idem</i>	
11. Nézetek létrehozása földrajzi és időbeli adatkészletekről	<i>idem</i>	
12. Hálózatelemzés - bevezető fogalmak. Grafikon. Csomópont. Fokozat. Központosság.	<i>idem</i>	
13. Társadalmi hálózatok/interakciók/együtműködések nézetek létrehozása. Hasonlósági együtthatók.	<i>idem</i>	
14. Szövegbányászat - természetes nyelvi feldolgozási technikák. Tudásfelfedezés és reprezentáció	<i>idem</i>	
Könyvészet:		
<ol style="list-style-type: none"> Riley, David D., and Kenny A. Hunt. Computational thinking for the modern problem solver. CRC Press, 2014. Foreman, John W. Data Smart: Using Data Science to Transform Information into Insight. John Wiley & Sons, 2013. Ferrara, Emilio, Pasquale De Meo, Giacomo Fiumara, and Robert Baumgartner. "Web data extraction, applications and techniques: A survey." Knowledge-Based Systems 70 (2014): 301-323 Ignatow, Gabe, and Rada Mihalcea. Text mining: A guidebook for the social sciences. Sage Publications, 2016. Social media importing via APIs and graph visualisation/analysis http://nodexl.codeplex.com/ Keyling, Till; Jünger, Jakob Facepager. An application for generic data retrieval through APIs. (2013). 		

7. Using OpenRefine https://www.packtpub.com/sites/default/files/9781783289080_Chapter_02.pdf
8. Gray, Jonathan, Lucy Chambers, and Liliana Bounegru. The data journalism handbook. " O'Reilly Media, Inc.", 2012. - datajournalismhandbook.org

8.2 Szeminárium / Labor	Didaktikai módszerek	Megjegyzések
1. Adatbányászat - Web Scraping	esettanulmányok	
2. Adatbányászat - Web Scraping	esettanulmányok, gyakorlatok	
3. Adatfeldolgozás/tisztítás - MS	bemutató, gyakorlatok	
4. Adatfeldolgozás/tisztítás - MS	gyakorlás	
5. Relációs adatok és hálózatelemzés - NodeXL	gyakorlás	
6. Adatvizualizáció – Tableau Public	gyakorlás	
7. Automatikus szövegfeldolgozás - KH Coder	gyakorlás	

Könyvészet:

1. Multi-level extraction https://www.youtube.com/watch?v=iFbA3-q_ZQk
2. Advanced extraction and online pre-mades https://www.youtube.com/watch?v=wOPuhYLg_6c
3. Navigating multi-level navigation to scrape all items in a dynamic site <http://webscraper.io/>
4. Import.io – fast linear scraping and API generation <https://import.io/>
5. Using Excel and add-ins (ASAP Utilities) <http://www.asap-utilities.com/>
6. Tableau Public <https://public.tableau.com>
7. NodeXL Excel add-in extension <http://nodexl.codeplex.com/>

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

- - Napjainkban egyre több vállalat érdeklődik az online piackutatás iránt, vagy nagy mennyiségű adatot gyűjt az online vállalkozások indításához. Az automatizált online adatgyűjtés is komoly alapja lehet az adatújságírásnak.

10. Értékelés

Tevékenység típusa	10.1 Értékelési kritériumok	10.2 Értékelési módszerek	10.3 Aránya a végső jegyben
10.4 Előadás	Alapfogalmak ismerete	Dolgozat	60%
10.5 Szeminárium / Labor	Szemináriumi témák előkészítése Záró projekt - legalább 10 000 rekordból álló adatbázis kinyerése, feldolgozása és vizualizációja.	5 szemináriumi tevékenység, amelyek egy záró projektet eredményeznek, amelyet a tanfolyam során ellenőriznek.	40%

10.6 A teljesítmény minimumkövetelményei

- A kurzus végén a hallgatók képesek lesznek arra, hogy a közösségi média API-k lekérdezésére alkalmas technikákat vagy eszközöket használjanak, hogy nagy mennyiségű adatot nyerjenek ki az online környezetből. A hallgatók minimális szinten képesek lesznek az adatbázisokkal való

működésre, ismerik az alapvető szöveges adattisztítási műveleteket, és képesek lesznek különböző típusú adatvizualizációk és értelmezések elvégzésére.

Kitöltés dátuma

2022.05.16.

Előadás felelőse

LECT. DR. Péter Árpád

Szeminárium felelőse



Az intézeti jóváhagyás dátuma

2022.05.30.

Intézetigazgató

CONF. DR. Radu Meza

