

FIȘA DISCIPLINEI

1. Date despre program

1.1 Instituția de învățământ superior	UNIVERSITATEA BABEȘ-BOLYAI
1.2 Facultatea	FACULTATEA DE ȘTIINȚE POLITICE, ADMINISTRATIVE ȘI ALE COMUNICĂRII
1.3 Departamentul	DEPARTAMENTUL DE JURNALISM
1.4 Domeniul de studii	ȘTIINȚE ALE COMUNICĂRII
1.5 Ciclul de studii	LICENȚĂ
1.6 Programul de studiu / Calificarea	MEDIA DIGITALĂ

2. Date despre disciplină

2.1 Denumirea disciplinei	ANALIZA DATELOR DIGITALE						
2.2 Titularul activităților de curs	RADU MEZA						
2.3 Titularul activităților de seminar	FLAVIA ȚĂRAN						
2.4 Anul de studiu	2	2.5 Semestrul	I	2.6. Tipul de evaluare	E	2.7 Regimul disciplinei	OB

3. Timpul total estimat (ore pe semestru al activităților didactice)

3.1 Număr de ore pe săptămână	3	Din care: 3.2 curs	2	3.3 seminar/laborator	1
3.4 Total ore din planul de învățământ	42	Din care: 3.5 curs	28	3.6 seminar/laborator	14
Distribuția fondului de timp:					ore
Studiul după manual, suport de curs, bibliografie și notițe					10
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					20
Pregătire proiecte					24
Tutoriat					
Examinări					2
Alte activități:					
3.7 Total ore studiu individual		56			
3.8 Total ore pe semestru		98			
3.9 Numărul de credite		4			

4. Precondiții (acolo unde este cazul)

4.1 de curriculum	•
4.2 de competențe	•

5. Condiții (acolo unde este cazul)

5.1 De desfășurare a cursului	• Sală cu videoproiector
5.2 De desfășurare a seminarului/laboratorului	• Laborator cu calculatoare Software: Helium Scraper, Chrome Web Scraper, 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. Competențele specifice acumulate

Competențe profesionale	<p>C2.1 Identificarea și înțelegerea teoriilor cu privire la gestiunea informațiilor, comunicarea mediatică, sistemele media, dinamica acestora și la efectele comunicării mediatică</p> <p>C2.3 Utilizarea tehnologiilor adecvate în realizarea și comunicarea profesionalizată a diverselor produse de informare, cu respectarea dreptului publicului de a fi informat corect</p> <ul style="list-style-type: none"> • Abilitatea de a utiliza programe și servicii ce permit configurarea de scrapere/wrapere Web; • Abilitatea de a utiliza programe și servicii ce permit interogarea API-urilor unor servicii Web; • Abilitatea de a utiliza programe pentru curățarea și organizarea unor seturi mari de date <p>C2.4 Utilizarea unor criterii și metode standard de evaluare a calității și limitelor folosirii diverselor tehnologii în cadrul unor acte concrete de informare și comunicare</p> <p>C2.5 Realizarea de portofolii cu privire la dimensiunea mediatică a unui act de informare și/sau comunicare profesionalizată</p> <ul style="list-style-type: none"> • Abilitatea de a folosi și agrega seturi mari de date pentru a genera vizualizări; • Capacitatea de a analiza fenomene Web pe baza unor vizualizări generate folosind seturi de date.
Competențe transversale	<ul style="list-style-type: none"> •

7. Obiectivele disciplinei (reieșind din grila competențelor acumulate)

7.1 Obiectivul general al disciplinei	<ul style="list-style-type: none"> • Dezvoltarea abilității de a colecta și utiliza seturi mari de date din mediul online
7.2 Obiectivele specifice	<ul style="list-style-type: none"> ▪ Înțelegerea conceptelor de bază referitoare la Web, paginile Web statice și dinamice, respectiv funcționarea site-urilor Web; ▪ Înțelegerea modului de funcționare al scraperelor/wrapperelor Web și API-urilor; ▪ Înțelegerea tipurilor principale de date și modului de prelucrare aferent acestora; ▪ Înțelegerea principiilor de construire a vizualizărilor de date și adaptarea acestora la seturi mari de date; ▪ Dezvoltarea capacității de a configura un scraper/wrapper folosind aplicații open-source; ▪ Dezvoltarea capacității de a organiza și curăța un set mare de date; ▪ Dezvoltarea capacității de a genera vizualizări ale unor seturi mari de date.

8. Conținuturi

8.1 Curs	Metode de predare	Observații
1. Web-ul ca sursă de date. Noțiuni introductive privind Web Wrapping/Scraping/Crawling – istoric și definiții.	Expunere, Explicație, Demonstrație	
2. Date publice și date deschise. Web Wrapping/Scraping/Crawling – probleme legale.	Expunere, Explicație, Demonstrație	
3. Pașii colectării automate de date folosind Web scrapere. Șabloane HTML. Simularea navigării	Expunere, Explicație, Demonstrație	

4. Configurarea unor acțiuni automate. Paginație. Autentificare. Extragere multi-nivel	Expunere, Explicație, Demonstrație	
5. Tipuri de date și structuri de date. Operații posibile cu tipuri de date. Reprezentarea datelor. Conversia datelor.	Expunere, Explicație, Demonstrație	
6. Baze de date relaționale – noțiuni introductive. Tipuri de date.	Expunere, Explicație, Demonstrație	
7. Baze de date. Tabele. Relații. Interogări. SQL – noțiuni de bază.	Expunere, Explicație, Demonstrație	
8. Interogarea API-urilor serviciilor social media/generice	Expunere, Explicație, Demonstrație	
9. Principii și tehnici de curățare și management pentru seturile de date extrase de pe Web – operații cu stringuri, categorisire automată	Expunere, Explicație, Demonstrație	
10. Analiza cuvintelor cheie. Variabile, metrice și analize specifice pentru optimizare pentru motoare de căutare (SEO)	Expunere, Explicație, Demonstrație	
11. Generarea de vizualizări pe seturi de date geografice și temporale	Expunere, Explicație, Demonstrație	
12. Analiza rețelelor – noțiuni introductive. Graf. Nod. Muchie. Grad. Centralitate.	Expunere, Explicație, Demonstrație	
13. Generarea de vizualizări ale rețelelor sociale/de interacțiune/de co-ocurență. Coeficienți de similitudine.	Expunere, Explicație, Demonstrație	
14. Text mining – tehnici de procesare a limbajului natural. Knowledge discovery & representation	Expunere, Explicație, Demonstrație	

Bibliografie:

1. Riley, David D., and Kenny A. Hunt. Computational thinking for the modern problem solver. CRC Press, 2014.
2. Foreman, John W. Data Smart: Using Data Science to Transform Information into Insight. John Wiley & Sons, 2013.
3. 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
4. **Ignatow, Gabe, and Rada Mihalcea. Text mining: A guidebook for the social sciences. Sage Publications, 2016.**
5. Social media importing via APIs and graph visualisation/analysis <http://nodexl.codeplex.com/>
6. 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 Seminar / laborator	Metode de predare	Observații
1. Extragerea datelor - Web Scraping	Aplicație practică Discuție	
2. Extragerea datelor - Web Scraping	Aplicație practică Discuție	
3. Procesarea/curățarea datelor - MS	Aplicație practică	

Excel	Discuție	
4. Procesarea / curățarea datelor - MS Access	Aplicație practică Discuție	
5. Date relaționale și analiza rețelelor – NodeXL	Aplicație practică Discuție	
6. Vizualizarea datelor – Tableau Public	Aplicație practică Discuție	
7. Prelucrarea automată a textului – KH Coder	Aplicație practică Discuție	

Bibliografie

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. Coroborarea conținuturilor disciplinei cu așteptările reprezentanților comunității epistemice, asociațiilor profesionale și angajatori reprezentativi din domeniul aferent programului

- In momentul actual tot mai multe companii sunt interesate de cercetari de piata in mediul online sau de colectarea de cantitati mari de date in vederea pornirii unor afaceri in mediul online. De asemenea, colectarea automata a datelor din mediul online poate constitui o baza serioasa pentru jurnalismul de date.

10. Evaluare

Tip activitate	10.1 Criterii de evaluare	10.2 metode de evaluare	10.3 Pondere din nota finală
10.4 Curs	Cunoașterea conceptelor de bază	Examen final (tip grilă)	60%
10.5 Seminar/laborator	Pregătirea temelor de seminar Proiect final – extragerea unei baze de date de cel puțin 10.000 de înregistrări, prelucrarea și realizarea unei vizualizări pe baza acesteia	5 activități de seminar concretizate într-un proiect final, verificat pe parcurs	40%

10.6 Standard minim de performanță

La finalul cursului studentii vor putea sa foloseasca tehnici de scraping/ wrapping sau unelte de interogarea a API-urilor platormelor social media pentru a extrage cantitati mari de date din mediul online. De asemenea studentii vor putea opera cu baze de date la un nivel minim, vor cunoaște operațiuni de bază de curățare a datelor în format text și vor putea realiza diferite tipuri de vizualizări și interpretări ale datelor.

Data completării

Semnătura titularului de curs

Semnătura titularului de seminar

.....

.....

.....

Data avizării în departament

Semnătura directorului de departament

.....

.....