

Óbudai Egyetem Alba Regia Műszaki Kar					
Tantárgy neve és kódja: Szöveg- és webbányászat (ATVSW4DMNF)			Kreditérték: 4		
Nappali/Levelező tagozat 2025/26 tanév I. félév					
Szakok, melyeken a tárgyat oktatják:					
Tantárgyfelelős oktató:			Oktatók:	Molnár Géza Gábor	
Előtanulmányi feltételek: (kóddal)					
Heti óraszámok:	Előadás: 1	Tantermi gyak.:	Laborgyakorlat: 2	Konzultáció:	
Számonkérés módja (s,v,f):	Évközi jegy				
A tananyag					
Oktatási cél: A kurzus célja a szöveg- és webbányászat fogalmainak, használt módszereinek és azok alkalmazásainak megismertetése a hallgatókkal. A hallgató képes lesz a témát érintő tipikus problémák megértésére, a megoldáshoz szükséges eszköz és módszer kiválasztására, valamint a probléma megoldására.					
Tematika: Strukturált és nem strukturált adatok. A szöveg- és webbányászat alapfogalmai. Természetes nyelvű feldolgozás. Szöveg előfeldolgozása, lexikális, szintaktikai és szemantikai elemzése. Ontológiák. Tipikus szövegfeldolgozási feladatok (Információkinyerés. Dokumentumok osztályozása. csoportosítása, rangsorolása és összegzése) és gyakorlati alkalmazások (Vélemény elemzés, Hangulat elemzés, Spam detektálás, Fordítás). A web szerkezete, keresőmotorok, indexelési módszerek. Web scraping. Webtartalom, struktúra és használat elemzése. A webbányászat tipikus alkalmazásai (Ajánlórendszerek, Közösségi hálók, Logfájlok elemzése).					
Témakör					Óraszám
Előadások/Gyakorlatok:					
Szöveg- és webbányászati alapfogalmak, tipikus problémák és alkalmazások					1+2
Szöveg előfeldolgozása I. (szózsák modell)					1+2
Szöveg előfeldolgozása II. (tokenizálás)					1+2
Szöveg előfeldolgozása III. (szövegtisztítás)					1+2
Lexikális, szintaktikai és szemantikai elemzés.					1+2
Tipikus szövegfeldolgozási feladatok I. (információ kinyerése)					1+2
Tipikus szövegfeldolgozási feladatok II. (szövegek összehasonlítása, klaszterezés)					1+2
Gyakorlati alkalmazások I. (vélemény- és hangulatelemzés)					1+2
Gyakorlati alkalmazások II. (összegzés, fordítás, spam-detektálás)					1+2
Webtartalom, struktúra, és használat elemzése					1+2
Rektori szünet					
Ajánlórendszerek, Közösségi hálók					1+2
Számonkérés (ZH)					1+2
Pótlás/Javítás					1+2
Félévközi követelmények					
Zárthelyi dolgozat (13. hét) + házi feladatok (4. és 8. heti kiadással)					
13. hét	Elméleti teszt + gyakorlati feladatok megoldása gépnél az 1-12. hét anyagából				
A pótlás módja:	A 14. héten				
Aláírás feltétele:	Zárthelyi dolgozat megírása legalább 50%-os teljesítménnyel, a házi feladatok határidőre történő leadása				
A vizsga módja (írásbeli, szóbeli, teszt, stb):					

Irodalom:	
Kötelező:	A Moodle rendszerben közzétett anyagok
Ajánlott:	Feldman, R., & Sanger, J. (2007). The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge University Press. Bing Liu (2011). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer