

Tantárgy neve: Webbányászat	Kreditszáma: 3
Tanóra típusa: előadás+gyakorlat, száma: 2+1	
Számonkérés módja: kollokvium	
Előtanulmányi feltételek (<i>ha vannak</i>):	
Tantárgyleírás:	
<p>Bevezetés, fogalmak. A szövegbányászat fogalma, feladata, alkalmazási területei. Adatbáziskezelés, információ-visszakeresés, adatabányászat, nyelvtechnológia és szövegbányászat kapcsolata</p> <p>A szöveges információ előfeldolgozása, dokumentum reprezentálása. A vektortér modell. Indexelési technikák, hasonlósági mértékek.</p> <p>Információ-visszakereső rendszer architektúrája. Világháló. A világháló története. A világháló napjainkban. Mély web, sekély web. A Szemantikus világháló.</p> <p>Tartalomkeresés webdokumentumban. Webkeresőmotorok. PageRank módszer.</p> <p>Az információ-visszakereső módszerek, rendszerek relevanciahatékonyságának mérése. Szabványos tesztadatbázisok a mérésekben. Leighton-módszer.</p> <p>Az információ-kinyerés fogalma. Az információ-visszakeresés és információ-kinyerés összehasonlítása. Tulajdonnév felismerés.</p> <p>Szövegbányászati módszerek. Osztályozás, csoportosítás. Kivonatolás.</p> <p>Véleményelemzés, hangulatbányászat (Sentiment analysis, opinion mining). Az ajánlórendszerek.</p> <p>„BEHÁLÓZVA”: hálózatok működése, szerveződése, közösségi hálók.</p>	
A 3-5 legfontosabb kötelező irodalom:	
Tikk Domonkos: Szövegbányászat, Típotext, Budapest, 2007	
Ajánlott irodalmak:	
Dominich, S.: Mathematical Foundations of Information Retrieval; Ricardo Baeza-Yates, Berthier Ribeiro-Neto: Modern Information Retrieval;	
Tantárgyfelelős (<i>név, beosztás, tud. fokozat</i>): Dr. Góth Júlia PhD, docens	
Tantárgy oktatásába bevont oktató(k) , ha vannak (<i>név, beosztás, tud. fokozat</i>):	