

Cuprins

1. Introducere	9
2. Scurtă istorie a DH.....	16
3. Metode în DH	22
3.1. Colectarea și adnotarea de corpusuri.....	23
3.2. Trecerea textului în format digital – OCR-izare	31
3.3. Curățarea corpusurilor	33
3.4. Interogarea corpusurilor	34
3.5. Procesarea automată a corpusurilor.....	39
3.5.1. Tokenizare, stematizare și lematizare.....	40
3.5.2. Etichetare cu părți de vorbire	42
3.5.3. Cuvinte gramaticale.....	47
3.5.4. Analiza sintactică automată a textelor	48
3.5.5. Analiza semantică a textelor.....	62
3.6. Analiza propriu-zisă a corpusurilor.....	85
3.6.1. Analiza cantitativă a corpusurilor.....	85
3.6.2. Analiza sentimentelor	93
3.6.3. Extragerea automată a temelor	96
3.6.4. Recunoașterea automată a entităților	100
3.6.5. Sarcini abordate prin tehnici de învățare automată tradițională.....	101
3.6.6. Sarcini abordate prin tehnici de învățare profundă.....	110
3.7. Evaluarea și interpretarea rezultatelor	113
3.7.1. Metrice de evaluare a similarității între documente	113
3.7.2. Metrice de evaluare a învățării automate	114
3.7.3. Teste statistice	119
4. Instrumente și biblioteci	136
4.1. Instrumente de analiză de textelor.....	137
4.2. Instrumente pentru vizualizarea datelor și a rețelelor	142
4.3. Biblioteci pentru procesare de text și vizualizare.....	145
4.4. LLM-uri disponibile ca instrumente.....	146

5. Aplicații	148
5.1. Analiza computațională a discursurilor politice ale președinților României.....	149
5.2. Identificarea și clasificarea automată a postărilor care conțin învinovățirea victimelor abuzurilor sexuale	156
5.3. Diferențe generaționale în exprimarea pe platforme de socializare	165
5.4. Detectarea evoluției stilistice de la comunism la democrație în cazul operei publicistice a lui Solomon Marcus.....	171
5.5. Măsurarea schimbării semantice a cuvintelor	179
5.6. Identificarea automată a cuvintelor înrudite în limbile romanice	186
5.7. Prezicerea automată a accentului în limba română	193
5.8. Testarea abilităților creative ale modelelor de limbaj mari și compararea acestora cu abilitățile creative umane	195
5.9. Testarea abilităților modelelor de limbaj mari de a pastșa opere literare și compararea acestora cu abilitățile umane	200
5.10. Analiza contrastivă multimodală a meme-urilor de tip <i>brain rot</i> în italiană și română.....	205
6. Concluzii	214
7. Bibliografie	221