

Povzetek

Osrednji temi mojega diplomskega dela sta algoritma latentno semantično indeksiranje in PageRank. Podrobneje sem opisala vsa matematična orodja, ki jih uporabljata algoritma, predvsem potenčno metodo in singularni razcep. Najprej sem obravnavala delovanje algoritma LSI, načine, kako izboljšati učinkovitost, in postopke posodabljanja. V drugem delu sem se ukvarjala z osnovnim PageRank modelom in metodo za njegovo reševanje. Opisala sem parametre v PageRank modelu in občutljivost PageRank vektorja na spremembe. Delovanje obeh algoritmov sem preizkusila na primeru.

Math. Subj. Class. (2010): 65F15, 15A18, 68P20.

Ključne besede: Singularni razcep, model vektorskega prostora, latentno semantično indeksiranje, utežitev, posodabljanje, potenčna metoda, HITS algoritem, Markovske verige, PageRank, občutljivost.

Key words: Singular value decomposition, vector space model, latent semantic indexing, weighting, updating, power method, HITS algorithm, Markov chains, PageRank, sensitivity.

Literatura

- [1] Michael W. Berry, Murray Browne, *Understanding Search Engines: Mathematical Modeling and Text Retrieval*, Second edition, SIAM Philadelphia, 2005.
- [2] Carl D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000.
- [3] Bor Plestenjak, *Numerična linearna algebra*, skripta, verzija 13. julij 2010.
- [4] Amy N. Langville, Carl D. Meyer, *Deeper Inside PageRank*, Internet Mathematics, strani 335–380, 2005.
- [5] Amy N. Langville, Carl D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, 2006.
- [6] Amy N. Langville, Carl D. Meyer, *A Survey of Eigenvector Methods of Web Information Retrieval*, The SIAM Review, strani 135–161, 2005.
- [7] Ricardo Baeza-Yates, Emilio Davis, *Web Page Ranking using Link Attributes*, In the Thirteenth International World Wide Web Conference, strani 328–329, New York, 2004.
- [8] Michael W. Berry, Susan T. Dumais, Gavin W. O'Brien, *Using linear algebra for intelligent information retrieval*, The SIAM Review, strani 573–595, 1995.
- [9] Susan T. Dumais, *Improving the retrieval of information from external sources*, Behaviour Research Methods, Instruments & Computers, strani 229–236, 1991.
- [10] Pavel Moravec, *Testing dimension reduction methods for text retrieval*, Department of computer science, VŠB- Technical university of Ostrava, Proceedings of the Dateso 2005, strani 113–124, 2005.
- [11] Slavko Žitnik, *Inteligentno pridobivanje podatkov iz poljubnih virov*, Diplomsko delo, Univerza v Ljubljani, 2010.