1 PageRank Papers

The original Brin and Page papers are [37] and [38]. Some others on the topic of PageRank are [14], [29], [55], [125], [75], [76], [124], [28], [142], [114], [91], [85], [52], [79]. A good survey of web searching with PageRank summary is [13]. Modeling web surfing with the 'back button': [58].

Updating PageRank papers: [43], [116], [115], [97], [98].

2 HITS and its modified versions Papers

The original HITS paper is [93]. Others are [100], [55], [54], [26], [79], [46], [47], [7], [48], [41], [61], [62], [121], [42], [92], [49]. Good survey paper: [33]. HITS for bibliometrics: [36].

3 LSI Papers

The first LSI paper [56] was followed by [20], [21], [23], [53], [50], [140], [141], [101], [87], [138], [89], [90], [94], [82], [88], [31], [30].

4 Similarity Measures

Some new similarity measures proposed are: [86], [68], [121].

5 Stats and Structure of Web

Several papers study the structure of the web and statistics regarding the web: [39], [3], [2], [80], [24], [1], [4], [5], [25], [60], [32], [12], [120], [119].

6 Laplacian and Graph Theory

The eigenvalues of the Laplacian of an undirected graph provide a surprising amount of information about the structure of the associated matrix. Do such findings extend to directed graphs, like those underlying a Markov chain? See [113], [112], [63], [64].
7 Clustering

Clustering is a difficult and important issue for IR. See [136], [34], [135], [51], [50], [77], [17], [67], [22], [78], [63], [64], [111], [45], [133].

8 Markov Chains

Classic Markov chain papers are [66], [72], [44], [107], [108], [110], [129], [57], [73], [134], [106], [81], [16], [84], [83], [128].

9 Books

Useful books for IR researchers are [40], [109], [132], [137], [131], [130], [118], [15], [18], [19], [65], [71], [95], [139].

10 Miscellaneous

Some papers that are harder to classify: [27], [102], [123], [99] deal with relevance feedback and user-tailored queries. [104] and [50] combine search engine results into one ranked list and [105] post-processes major search engine results annotating retrieved documents with topic information [103]. Spider's guide to PageRank: [126]. Meta-search engines: [35]. TREC-5 conference: [74]. Medline data set: [6]. An ML info. retrieval model: [117]. IR and genetic algorithm: [122]. Inquisit 2 metasearch engine [70, 99]. Sampling pages uniformly from the Web [127]. Using maximum likelihood and training sets to detect “hidden” links between groups, such as for government intelligence applications: [96]. Using a training set to classify documents and incorporating text content and link information: [69]. Several papers by same authors on matrix perturbation and spectral decomposition: [11], [8], [9], [10].

References


[43] Steve Chien, Cynthia Dwork, Ravi Kumar, and D. Sivakumar. Towards exploiting link evolution.


