Ranking algorithms

http://static.googleusercontent.com/media/research.google.com/de//pubs/archive/41657.pdf - Data Fusion: Resolving Conflicts from Multiple Sources – AccuCopy alogithm – discover true values by considering accuracy of a copying between sources. Determine if sources are independent. Copy detection used. AccuCopy converges if source accuracy is ignored.

http://www.ksi.mff.cuni.cz/~pokorny/papers/IADIS-AP05.pdf - Page Content Rank: An approach to web content mining- propose a new Web Content Mining method of a page relevance ranking based on the page content exploration; focused only on exploring the content of pages (unlike PageRank and HITS); page content rank (PCR) - term extraction, parameters calculation, term classification, calculation of page relevance

https://papers.nips.cc/paper/4701-iterative-ranking-from-pair-wise-comparisons.pdf - Iterative Ranking from Pair-wise Comparisons – model independent algorithm; finds "scores" for each object of interest to understand the intensity of the preferences; random walk approach to ranking

https://arxiv.org/ftp/arxiv/papers/1208/1208.1926.pdf - Role of Ranking Algorithms for Information Retrieval

- Page Rank - considers the back link in deciding the rank score.
- Weighted Page Rank - modification of the original PageRank algorithm. WPR decides the rank score based on the popularity of the pages by taking into consideration the importance of both the inlinks and outlinks of the pages. Every out-link page is given a rank value based on its popularity. Popularity of a page is decided by observing its number of in links and out links.
- HITS - a link based algorithm. In HITS algorithm, ranking of the web page is decided by analyzing their textual contents against a given query
- Distance Rank Algorithm - The main goal of this ranking algorithm is computed on the basis of the shortest logarithmic distance between two pages and ranked according to them so that a page with smaller distance to assigned a higher rank.
- EigenRumor Algorithm
- TF-IDT – occurance of words in document taken into consideration

Criteria/Algorithms	PageRank	Weighted PageRank	HITS	Distance Rank	EigenRumor
Mining techniques	WSM	WSM	WSM & WCM	WSM	WCM
Working process	Computes values at index time and results are sorted on the priority of pages	Computes values at index time and results are sorted on the basis of page importance	'n' highly relevant pages are computed and find values on the fly	Calculating the minimum average distance between two pages and more pages	Use the adjacency matrix which is constructed from agent to object link not page to page
Parameters	Inbound links	Inbound and outbound links	Inbound and outbound links and content	Inbound links	Agent/Object
Complexity	O(log N)	<O(log N)	<O(log N)	O(log N)	<O(log N)
Limitations	Query independent	Query independent	Topic drift & efficiency problem	Needs to work along with PageRank	Used for blog ranking

WSM – Web Structure Mining

WCM – Web content Mining

Problem statements:

Given a list of claims for a particular event, produce a ranking of values by relevance. Event categories are taken into consideration.

Value types might be different (conflicts) – rank according to data types and event types

Some values may be taken from other sources (copied or referenced). Merge/fuse data for optimal results.