

Talk on “WEB PAGE Ranking using Domain Based Knowledge” IEEE paper

Akarsh Kashamshetty



Motivation

Numerical computation.

Google Page Rank.

Related Work

- Probability based ranking
- Graph based ranking
- Concept based ranking
- Machine learning based ranking algorithms

Key concepts

- Primary Web pages – Webpages whose keywords match with the user query and are ranked based on the primary quotient value.
- Secondary Web pages – Web pages whose keywords do not match completely with the user query but contents match with the user requirement.

$$R_1(W) = \left(\frac{1}{\sum PR(W)} + \frac{\sum S_w}{N_w} \right)$$

Equations

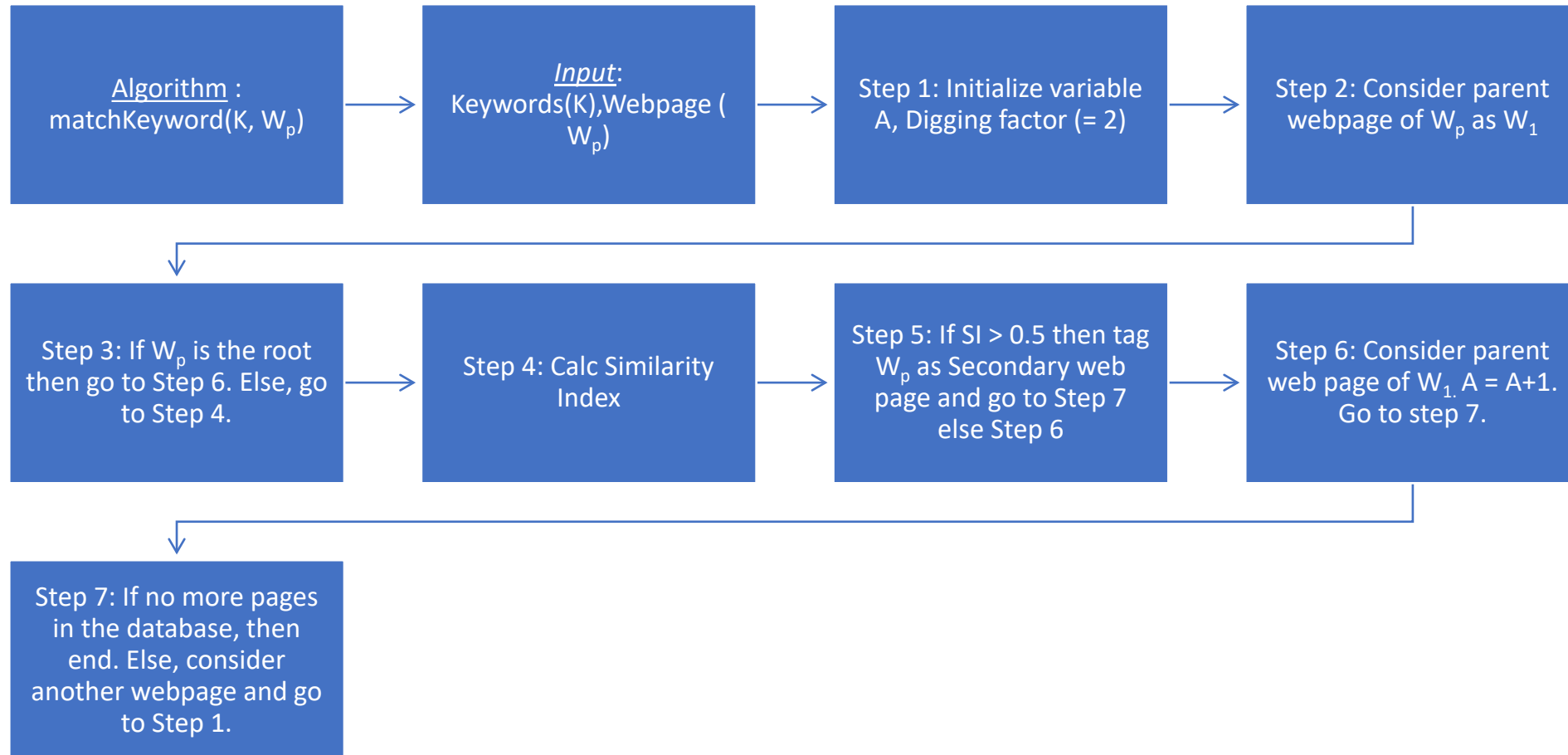
R_1 , R_2 - equations for calculating the primary quotient (α rank) for primary and secondary web pages.

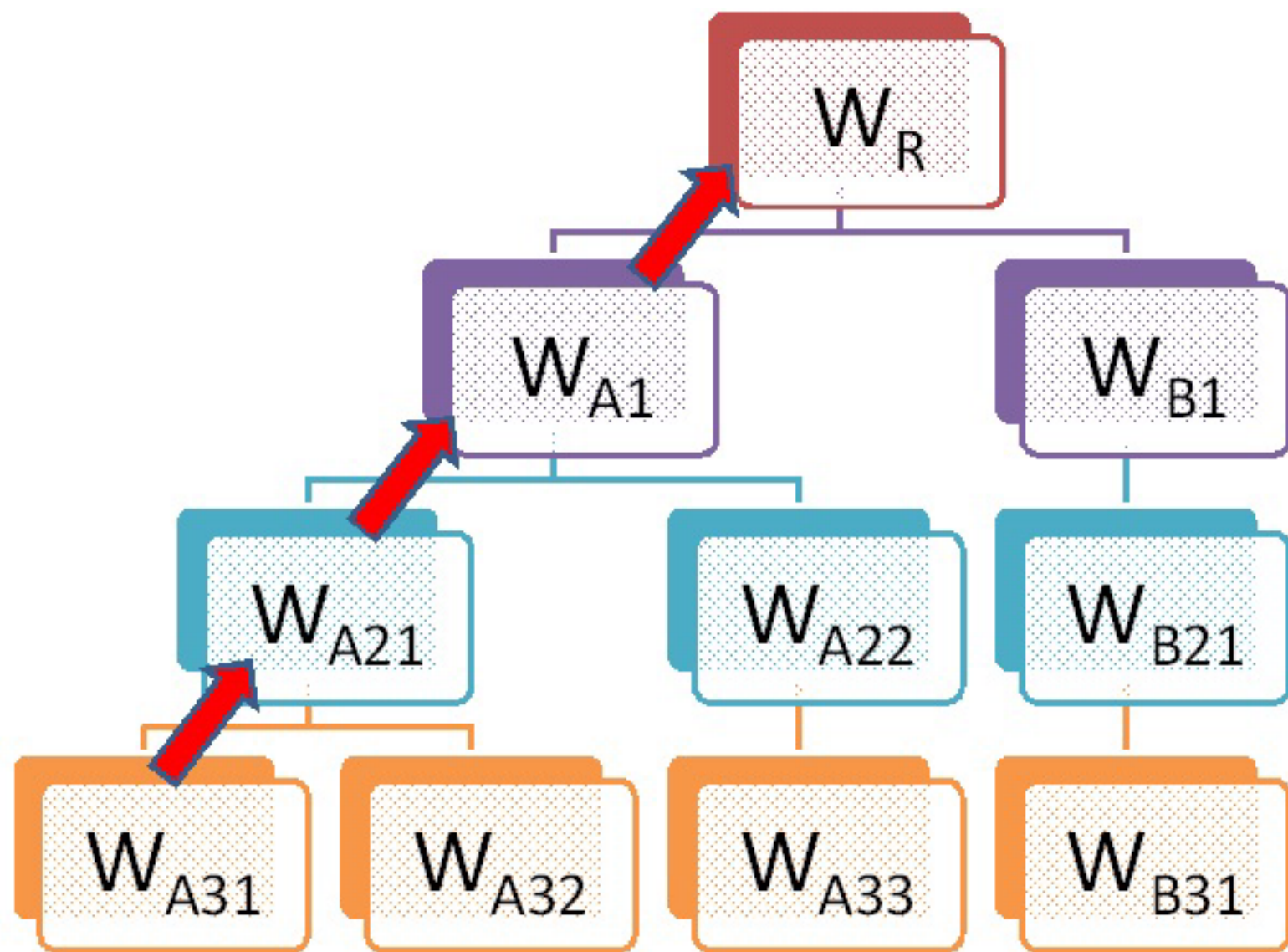
I_s – Similarity Index

$$I_s = \frac{(K_{FM} + 0.5) \times K_{0.5}}{K_N(W)}$$

$$R_2(W) = \frac{\left(\frac{1}{\sum PR(W_1)} + \frac{\sum S_w}{N_w} \right)}{D_F}$$

Secondary webpage selection





Case Study : K = “Multi Agent”

Details of web pages for user query

Web Page	Web Page(W)	Classification
WI	http://www.capexindia.com/IRL/index.htm	UNMATCHED
WM	http://www.capexindia.com/IRL/mission.htm	UNMATCHED
WV	http://www.capexindia.com/IRL/vision.htm	UNMATCHED
WR	http://www.capexindia.com/IRL/research.htm	MATCHED
WAI	http://www.capexindia.com/IRL/centre_for_AI.htm	UNMATCHED
WES	http://www.capexindia.com/IRL/centre_for_ES.htm	UNMATCHED
WCA	http://www.capexindia.com/IRL/centre_for_CA.htm	UNMATCHED
WMAS	http://www.capexindia.com/IRL/centre_for_MAS.htm	MATCHED
WWEB	http://www.capexindia.com/IRL/centre_for_web.htm	UNMATCHED
WTI	http://www.capexindia.com/IRL/centre_for_TI.htm	UNMATCHED
WT	http://www.capexindia.com/IRL/team.htm	UNMATCHED
WPJ	http://www.capexindia.com/IRL/published_journal.htm	MATCHED
WPBC	http://www.capexindia.com/IRL/book_chapter.htm	MATCHED
WPC	http://www.capexindia.com/IRL/published_conference.htm	MATCHED
WCP	http://www.capexindia.com/IRL/communicated_papers.htm	UNMATCHED
WOGP	http://www.capexindia.com/IRL/on_going_papers.htm	MATCHED
WMEM	http://www.capexindia.com/IRL/member.htm	UNMATCHED
WJP	http://www.capexindia.com/IRL/journal_participation.htm	UNMATCHED
WCPART	http://www.capexindia.com/IRL/conference_participation.htm	UNMATCHED
WPROJ	http://www.capexindia.com/IRL/projects.htm	MATCHED
WCON	http://www.capexindia.com/IRL/contact.htm	UNMATCHED

Ranks of Primary and Secondary

Web page (W)	$\Sigma(\text{PR}(\text{W}_i))$	ΣS_w	N_w	$R_1 (W)$ value
http://www.capexindia.com/IRL/research.htm	04	14	04	3.5
http://www.capexindia.com/IRL/centre_for_MAS.htm	04	32	03	10.91
http://www.capexindia.com/IRL/published_journal.htm	04	26	05	5.45
http://www.capexindia.com/IRL/book_chapter.htm	04	07	03	2.58
http://www.capexindia.com/IRL/on_going_papers.htm	04	07	03	2.58
http://www.capexindia.com/IRL/projects.htm	04	09	02	4.75

Web page	Σ ($\text{PR}(\text{W}_i)$)	Σ S_w	N_w	Digging Factor (D_F)	R_2 (W) value
http://www.capexindia.com/IRL/index.htm	04	24	10	01	2.65
http://www.capexindia.com/IRL/centre_for_AI.htm	04	21	02	03	3.58
http://www.capexindia.com/IRL/centre_for_web.htm	04	26	02	03	4.41
http://www.capexindia.com/IRL/communicated_papers.htm	04	32	03	03	3.63

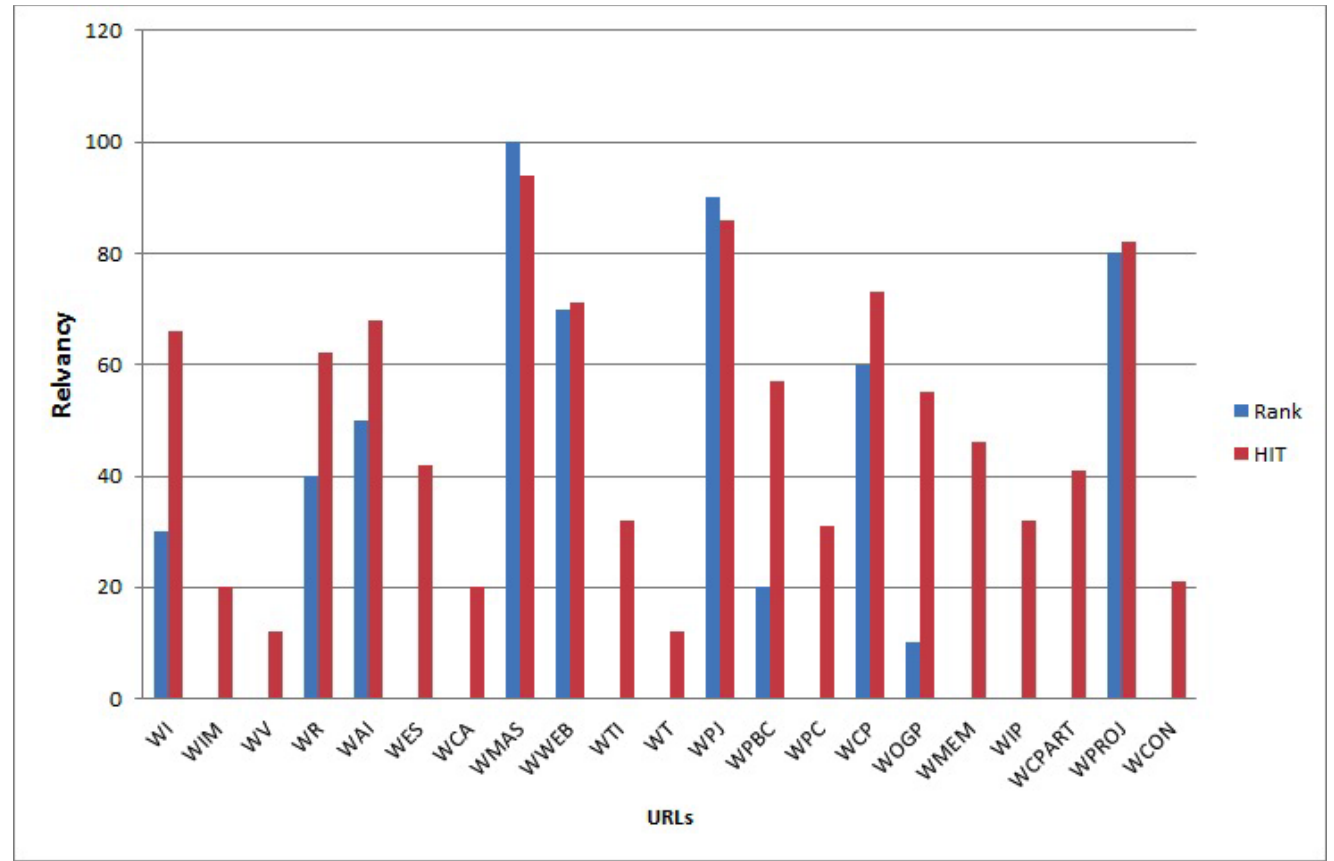
Rank of the Webpages

- Altogether(primary and secondary)

Web pages	Web pages (URL)	Quotient Value	Rank
WR	http://www.capexindia.com/IRL/research.htm	3.5	07
WMAS	http://www.capexindia.com/IRL/centre_for_MAS.htm	10.91	01
WPJ	http://www.capexindia.com/IRL/published_journal.htm	5.45	02
WPBC	http://www.capexindia.com/IRL/book_chapter.htm	2.58	09
WOGP	http://www.capexindia.com/IRL/on_going_papers.htm	2.58	10
WPROJ	http://www.capexindia.com/IRL/projects.htm	4.75	03
WI	http://www.capexindia.com/IRL/index.htm	2.65	08
WAI	http://www.capexindia.com/IRL/centre_for_AI.htm	3.58	06
WWEB	http://www.capexindia.com/IRL/centre_for_web.htm	4.41	04
WCP	http://www.capexindia.com/IRL/communicated_papers.htm	3.63	05

Approach and ideal situation

- HIT and Rank
- Real time user favorite URLs and ranked URLs are matched.



References

- S. K. Guha, A. Kundu and R. Dattagupta, "Web page ranking using domain based knowledge," *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 1291-1297, doi: 10.1109/ICACCI.2015.7275791.
- F. Provost, P. Domingos, "Tree Induction for Probability-Based Ranking", *Machine Learning*, Volume: 52, Issue: 3, 2003, pp. 199-215.
- R. Mihalcea, "Graph-based ranking algorithms for sentence extraction, applied for text summarization", the ACL 2004 on Interactive poster and demonstration sessions (ACLDemo'04), July, 2004