The PageRank Citation Ranking

---- L. Page.

Summary

The Intentions of this paper are

- Rating webpages objectively and mechanically by using page rank.
- Computing page rank effectively by measuring human interest based on graph of web and applying to search and navigation.
- Taking advantage of Link structure of the web to produce a global importance ranking of every web page.
- Differentiating web pages with an approximation of overall relative importance of web page.
- Mathematical description of page rank, computing page rank, using page rank for searching and browsing.

Page ranking provides sophisticated methods for citation counting and considers importance obtained by link structure. Page rank is helpful in user navigation and returning effective search results for a keyword search.

The Main Ideas proposed in the paper are

- The Link structure of the Web: Forward Links, Back Links.
- Page Rank: A page has high rank if sum of ranks of its backlinks is high.
- Mathematical calculation of page rank using Eigen vectors and rank of back link pages.
- Rank Sink: Two web pages that pointing to each other but no other page forms a loop and any rank cannot be distributed which is known as rank sink.
- Rank Source: Calculating page rank of rank sink using a decay factor.
- Random Surfer Model: Page rank is calculated based on probability distribution of random walk where the surfer keeps on clicking on successive links randomly and jumps to a random page if he gets bored or entered a loop.
- Computation of page ranks based on vectors.
- Dangling Links: Links that point to any page with no outgoing links. We remove them until the page ranks are calculated.

- Implementing Page rank:
  - Web crawler creates the index of links and store all URL's in a database and assign a unique ID to each URL.
  - Sort link structure with parent ID.
  - Remove Dangling links.
  - Initial Assignment of Ranks.
  - Memory allocation for weights of each page.
  - Link database is sorted such that it can be stored on disk.
  - After weight converge, add dangling links back.
  - Re-compute page ranks.
- Page rank is linearly scalable because Web is expanding like a graph.
- Title Search using Page rank: Searches the title of all web pages for given query and then sorts based on page rank. This results in high precision and high quality.
- Rank Merging: Combining page rank for more specific searches where recall is more important. Google uses rank merging.
- Explained page ranking for common case approach which returns a commonly used commercial site that had little information rather than providing pages in detail.
- Personalized Page Ranking: Ranking pages based on personal home page and search engine.
- Page ranking based on commercial Interest and buying advertisements.
- Applications of Page Ranking:
- He gave basic definitions of
  - Estimating Web traffic: High usage Vs High Rank
  - Back Link Prediction: Effective way for crawling order and predicting citation counts.
  - User Navigation: The page rank proxy used to annotate each link the user visits with its page rank which helps in deciding links which are more likely to be interested.
  - Sorting Back links such that the best back links are displayed first.

The final outcome of paper is

- Global ranking of all web pages based only in their location in web's graph structure.

- Generating personalized page ranks to view web from particular perspective.

- Ordering search results that are more important and giving importance to back links from more significant pages.

- Structure of web page is helpful for various information retrieval tasks.

Submitted By

Lakshman Madhav Kollipara

SID- 932655504

Submitted To

Arash Termehchy

CS540- Database Management Systems