ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202310.01351

Using Search Engines to Analyze Communication Effects: A Research Postprint

Authors: Zhang Shouxian, Xu Zhiqiang, Li Manjiang

Date: 2023-10-08T00:00:00+00:00

Abstract

Utilizing news items retrieved from search engines to analyze and derive news dissemination effects and paths can serve as an important parameter for news quality and influence analysis. This paper discusses in detail the feasibility and general implementation steps, and analyzes its advantages and disadvantages.

Full Text

Research on Implementing Communication Effect Analysis Using Search Engines

Abstract: Using news items retrieved from search engines to analyze news communication effects and communication paths can serve as an important parameter for news quality and influence analysis. This paper discusses in detail the feasibility and general implementation steps, and analyzes its advantages and disadvantages.

Keywords: Communication Effect; Search Engine; Big Data

CLC Number: TP393.0

Document Code: A

Article ID: 1671-0134(2019)10-117-03

DOI: 10.19483/j.cnki.11-4653/n.2019.10.036

Authors: Zhang Shouxian[1], Xu Zhiqiang[2], Li Manjiang[3]

As a medium for news dissemination, new media has become the primary channel for news distribution, revolutionizing the traditional media landscape and causing severe market contraction for conventional outlets. New media' s interactive, timely, convenient, diverse, and personalized characteristics provide

readers with a superior experience while enabling news producers to better understand their audience. Communication effect analysis, which informs decisions about content, format, and timing, differs fundamentally from traditional media practices and has become a crucial indicator of news production quality and a key component of news big data analytics platforms. Such analysis can reveal readership, forwarding rates, geographic and demographic distribution, development trends, and media influence.

Currently, data for communication effect analysis is obtained through scheduled scraping of websites, apps, Weibo, WeChat, and other platforms. While this yields comprehensive and accurate data for analyzing numerous news items simultaneously, it suffers from significant drawbacks: massive storage requirements, high bandwidth consumption, low effective data ratios, and generation of substantial invalid internet traffic (scraped data often goes unused). The self-scraping process is extremely tedious, requiring tailored analysis for each website with different parsing methods. Frequent website updates create substantial workloads, and technical defenses against crawlers—such as those employed by Sina Weibo—make large-scale data scraping difficult. Consequently, this approach is unsuitable for ordinary news organizations seeking to build their own analysis platforms.

This paper proposes using search engines for news communication effect analysis, which can rapidly obtain required data with minimal storage overhead, making it ideal for typical news organizations. Leveraging search engine data eliminates cumbersome scraping efforts and associated costs. By using news titles as search keywords, results provide the total volume of news items, with each entry containing essential information such as media name, publication time, and original link. Accessing these original links yields detailed data including news sources, precise publication times, and in some cases, readership, forwarding, and comment metrics—encompassing all necessary data for general communication effect analysis.

1. Feasibility

The big data required for communication effect analysis already exists within major search engine companies, eliminating the need for independent scraping. Search engine interfaces can extract necessary data, and leveraging their vast stored datasets avoids redundant web-wide crawling. Directly utilizing search engine data circumvents the challenges of customized parsing for each website, frequent updates, and anti-crawling defenses, significantly reducing costs while providing all essential metrics for communication effect analysis.

2. Implementation Steps

Although programming effort for search engine-based analysis is relatively small, the following steps are required:

(1) Use the news title as a keyword to query the search engine.

(2) Analyze search engine results to obtain the total forwarding count.

(3) Utilize search engine pagination to compile a complete list of forwarding websites.

(4) Remove duplicate entries from the list.

(5) Analyze media names, publication times, and original links in the list.

(6) Open each news item' s original link to extract detailed information including specific title, news source, precise publication time, forwarding volume, readership, and other metrics.

(7) Based on media names, news sources, and publication times from each original link, generate communication paths and trend analyses.

The subsequent analysis and visualization components are relatively straightforward.

This method can effectively implement news communication effect analysis while achieving order-of-magnitude cost savings compared to traditional approaches, particularly suitable for news organizations producing fewer original daily stories. It significantly reduces storage and bandwidth requirements by retrieving only key news items and their publishing media, eliminating invalid internet traffic and conserving bandwidth. The analysis results can resolve evaluation challenges for new media content.

However, the method's accuracy depends on search engines' data coverage. Most search engines only index websites, lacking data from apps, Weibo, WeChat, and other news dissemination channels—representing a notable limitation, though using multiple search engines simultaneously can partially compensate. Nevertheless, website-based dissemination largely reflects news influence, generally not affecting comparative analysis of different news items' communication effects. For news primarily disseminated through Weibo, WeChat, or apps, analysis results may show significant deviation.

Thus, using search engines for communication effect analysis offers an economical and viable alternative solution. Of course, comprehensive news big data analytics platforms encompass more than just communication effect analysis, including public opinion monitoring, trending topics, sentiment analysis, user profiling, demographic analysis, and geographic analysis as additional reference indicators.

[Figure 1: see original paper] Search Engine Effect

[Figure 2: see original paper] Analyzed Communication Path

[Figure 3: see original paper] Analyzed Communication Trend

Funding: This work is supported by the Weifang Science and Technology Development Plan Project (Project No. 2019ZJ1162).

References

[1] Xu Zhiqiang, Zhang Shouxian, Li Manjiang. Research on Cloud Computing Big Data Platforms[J]. China Media Technology, 2019(9).

[2] Chen Si. Review of Practice and Theoretical Research on Big Data-Based Communication Effect Analysis[J]. News Dissemination, 2013(7).

[3] Jiang Yuxian. A Brief Discussion on News Communication Effect Research[J]. Zhiyin Inspirational, 2017(10).

Author Affiliations: 1. Peninsula Metropolitan Daily 2,3. Weifang Beida Jade Bird Huaguang Phototypesetting Co., Ltd.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.