Journal of Knowledge-Research Studies

Bahari Varzaneh, Hosein (2022). Comparing the Performance of Information Retrieval of Semantic and Keyword Search Engines Based on Phrase Search. *Journal of Knowledge-Research Studies*, 1 (2), 102-117.

DOI: 10.22034/JKRS.2022.51351.1014
URL: https://jkrs.tabrizu.ac.ir/article_15085.html

The paper is an open access and licensed under the Creative Commons CC BY NC license.

Comparing the Performance of Information Retrieval of Semantic and Keyword Search Engines Based on Phrase Search

Hosein Bahari Varzane*[1]

Received: April, 21, 2022; **Accepted: June, 17, 2022**

Abstract

Purpose: The aim of the research is to compare the performance of Information Retrieval of semantic and keyword search engines based on phrase search (simple & complex).

Methodology: The present applied and semi-experimental research community includes all active search engines on the web. Research samples were selected based on stratified random sampling and purposive sampling. The data collection tool of two researcher-made checklists includes ten simple and complex phrase queries.

Findings: Bing and Cluuz (with similar precision of 53%), DuckDuckGo, and Yahoo were the most accurate in searching for simple phrases, respectively. Bing, DuckDuckGo, Yahoo, and Cluuz were the most accurate in their search for complex terms, respectively. In general, Bing, DuckDuckGo, Cluuz and Yahoo have the highest precision, respectively. Also, the average total precision of keyword search engines is more heightened than semantic search engines.

Conclusion: The Bing keyword search engine performs better than the other three semantic search engines and other keywords. Semantic search engines claim to have more capabilities in retrieving relevant information than keyword search engines. But in this study, it was found that Cluuz and DuckDuckGo do not excel in search terms over keyword search engines. These tools did not perform as well as semantic web search engines, and it seems that they have a long way to go to become real semantic search engines. And to achieve this, it is necessary to use the facilities, tools, modules, and emerging technologies of the new age, such as machine learning, deep learning, combining these modules with pervasive techniques, data mining, etc.

Value: So far, not been compared the phrase search performance in the sample semantic and keyword search engines. And in this regard, the researcher has tried to achieve an actual result with an exact Survey.

Keywords: ***Information Retrieval, Phrase Search, Performance, Keyword Search Engines, Semantic Search Engines***

1. Master, Department of Knowledge and Information Science, Faculty of Education and Psychology, University of Isfahan, Isfahan, Iran (Corresponding Author) bahari_2062@yahoo.com

Extended Abstract

Introduction

The emergence of the web and the pervasiveness of information production among the general public has made the field of information retrieval an important issue (Nadi Ravandi & Hariri, 1394, 17-18). However, successful data retrieval from the web has its limitations. Like a. Content without proper structure for displaying information, b. Ambiguity in information due to the weak link between them, c. Inability to understand the information machines provide due to the lack of a global format, and d. The lack of automatic transfer of information leads to the retrieval of many irrelevant records (Sudeepthi et al., 2012). Although there are many search engines today, the main challenge of these scams is the inability to retrieve information intelligently (Gayar, Mekky, & Atwan, 2015). The information retrieved by these search engines is associated with high retrieval and low precision, with or without retrieval (Fatima, Luca, & Wilson, 2014). Semantic web technology is a solution to this problem. Semantic Web Achievement is Semantic Search Engines (S.S.E.); these tools claim to have advantages over keyword search engines: a. unlike keyword search engines (K.S.E.) that use the page ranking algorithm; S.S.E use ontology (Malve, & Chawan, 2015); B. They work according to a semantic approach (Sahu, Mahapatra & Balabantaray, 2015) and analyze the user's desire and identify the real and hidden meaning of the user's question; J. using logical reasoning, they retrieve more accurate results (Dorri, 2014).

Journal of Knowledge-Research Studies

Journal of Knowledge-Research Studies (JKRS)

Vol 1

Issue 2

Spring 2022

Purpose

Internet users do not know much about the performance of S.S.E and keywords. Also, it is not clear how these S.S.E and keywords perform in retrieving information for phrase queries (simple and complex). The purpose of this research is to identify a successful search engine and compare the performance of these two tools in this regard.

Methodology

The statistical population of this applied and quasi-experimental research includes all active search engines on the web. First, based on a random sampling of search engines, it was divided into two categories: S.S.E and keyword. Then, in these two categories of search engines, based on the superiority and use of search engines in previous research works and their popularity with users, they were selected from S.S.E, DuckDuckGo, Cluuz, and K.S.E, Yahoo, and Bing. The data collection tool has created two researcher checklists that include 10 query samples for phrase search (simple and complex). The relevance of the first 20 results retrieved for each query was examined, and the barrier of these tools was calculated using the precision coefficient. Finally, for each search engine, an average of the total obstacles was obtained, based on which their performance values were compared.

Findings

Findings obtained to achieve the objectives of the research, according to Figure 1, show that Bing and Cluuz search engines with the same precision of 53% achieved the highest precision in searching for type 1 expressions, which are considered simple expressions. Next to them is the DuckDuckGo search engine, and Yahoo's search engine showed the least precision in this study, with a minimal difference after DuckDuckGo. In addition, this chart shows that the Bing search engine has shown the highest level of precision in searching for Type 2 expressions, which are considered complex expressions. Then there are DuckDuckGo, Yahoo, and Cluuz search engines, respectively. Overall, the meanings obtained from the phrase search in these search engines showed that Bing, DuckDuckGo, Cluuz, and Yahoo retrieved the most relevant records, respectively.

Journal of Knowledge-Research Studies (JKRS)

Vol 1

Issue 2

Spring 2022

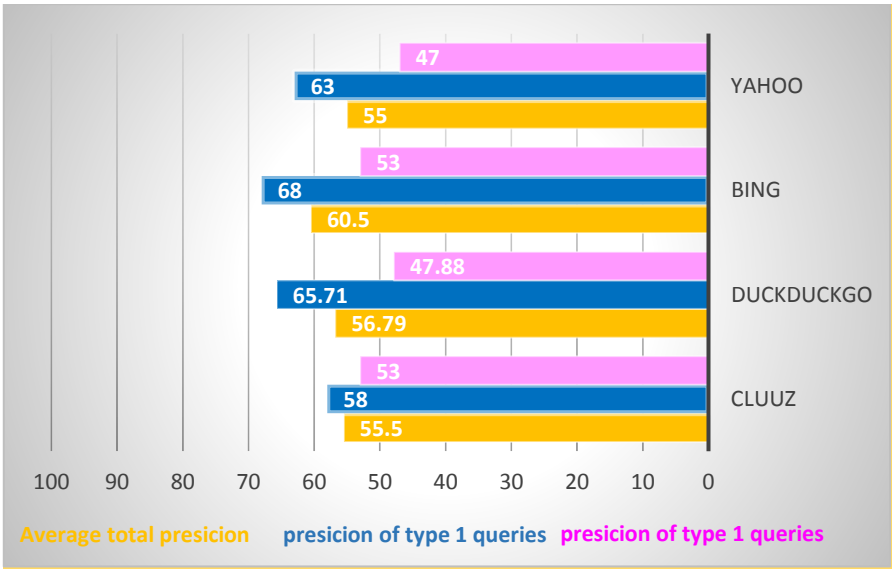


Figure 1. The average precision of search engines

The average total precision of S.S.E and K.S.E. is 57.75% and 56.145%, respectively. Based on this, the sample S.S.E has higher average precision than the average total accuracy obtained by K.S.E.

Conclusion

In simple term search, the Bing search engine performed best. This search engine can make decisions and provide immediate answers with multilingual support, and the type of results presented in it is free text and links (Negi & Kumar, 2014). Its unique feature is the "community" that allows learning and sharing resources with other people who use Bing (Dutta & Bansal, 2016). After Bing, DuckDuckGo had the best performance. DuckDuckGo goes to retrieve information with instant answers (Nandy et al., 2016) and results in categorized summaries. These results are based on many sources, including Yahoo, Wikipedia, Wolfram Alpha, and its crawler web. It has multilingual support, and the primary approach in it is cluster search and natural language processing (Negi & Kumar, 2014). Next to DuckDuckGo is a cluster whose results include cluster diagrams, images extracted from recovered results, link tags, and semantic graphs. It is possible to display the results related to the user search on the left side of the results page. Yahoo also showed the weakest performance by recovering the lowest number of related records. Yahoo, a type of search engine (Davarpanah, 1390, 96), has a hierarchical structure, and its main content is focused on financial news (Dutta and Bansal, 2016). The search results are provided along with a brief description of the retrieved websites (Davarpanah, 1390, 96), and there is a possibility of selective search based on it (Rathee, et al., 2013). In searching for Type 2 queries, which are complex, Bing performed best. It is followed by DuckDuckGo, Yahoo, and Cluuz, respectively. In searching for the two types of sample phrases, the Bing search engine performed better than Yahoo, and the dock search engine performed better than Cluuz. In general, the meanings obtained from the phrase search in these search engines showed that Bing was more powerful than the two S.S.E DuckDuckGo, Cluuz, and Yahoo search engines. This indicates that DuckDuckGo and Cluuz 's S.S.E, although they use semantic web standards, have a weak expression compared to the Bing search engine. The Bing search engine appears to be meaningful. Therefore, it is recommended that users visit the Bing search engine for a phrase search.

The findings from the present study are consistent with the results of Singh and Sharan (2013), Ahmad Khan, Sangroha, & Rahman (2014), and Nagi, & Kumar (2014). But the results of Mala & Lobiyal Research (2016) are inconsistent with the results of the present study. The number of inquiries selected for performance evaluation appears to be the number of records reviewed by the queries, the searches in search engines, the new features and tools used by the search engines, and most importantly busy time and time. And the web-based non–webmaster engines perform strong

Journal of Knowledge-Research Studies

Journal of Knowledge-Research Studies (JKRS)

Vol 1

Issue 2

Spring 2022

or poor at different times. The present study is also consistent with some of the previous research and some of them inconsistent.

In a general conclusion, it can be said that although S.S.E claims to be a: Understanding users' queries; B. Realizing what the meaning of the user's inquiry and purpose is; c. Responding to long and complex queries; d. Adapting the concepts, generalization, text mining on web pages; e. Comparing links by RDF and provide graph patterns; f. Benefiting from the rules of inference, logic, and natural language system (Ramachandran, A., & Sujatha, 2011), But in this study, it was found that the DuckDuckGo and Cluuz, when searching for a phrase, are not significantly superior to K.S.E, and only a few hundredths of a few hundred are careful. Therefore, these search engines did not show an acceptable performance as web search engines. They seem to have a long path to becoming real S.S.E. To achieve this, it is necessary to use the facilities, tools, modules, and emerging technologies of the new age from machine learning, deep learning, and integrating these modules with text techniques, data mining, and so on.

Value

So far, in the S.S.E and k.S.E., the data recovery performance has not been reviewed and compared through a phrase search. In this regard, the researcher has tried to achieve an actual result with a careful examination to introduce the powerful search engine in this type of search to users.

Journal of Knowledge-Research Studies (JKRS)

Vol 1

Issue 2

Spring 2022

Extended Abstract References

Khan, J.A., Sangroha, D., Ahmad, M., & Rahman, M.T. (2014). A performance evaluation of semantic based search engines and keyword based search engines. 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 168-173.

Aldabbagh, M. S. M. (2012). An Approach for Evaluation of Semantic Performance of Search Engines: Google, Yahoo, Msn and Hakia. *Journal of Education Science*, 25,124-130.

Aliyu, F. M., & Yahaya, Y. I. (2021). An Investigation of the Accuracy of Knowledge Graph-base Search Engines: Google knowledge Graph, Bing Satori and Wolfram Alpha. *International Journal of Scientific & Engineering Research*, 12(1), 11-15.

Azizan, A., Abu Bakar, Z., Abd Rahman, N., Masrom, S. & Khairuddin, N. (2018). A Comparative Evaluation of Search Engines on Finding Specific Domain Information on the Web. *International Journal of Engineering & Technology*, 7 (4), 1-4.

Choudhury, N. (2014). World Wide Web and Its Journey from Web 1.0 toWeb 4.0. Nupur Choudhury/ (IJCSIT). *International Journal of Computer Science and Information Technologies*,5(6),8096-80100.

Davarpanah, M.R. (2011). *Scientific Information Seeking in Printed and Electronic*. Tehran: Dabizash.

Dorri, R. (2015). Comparison and Evaluation of Semantic Search Engines. *Iranian Journal of Information processing and Management*, 30 (2) ,467-490. [In Persian].

Dutta, M., & Bansal, K. L. (2016). A Review Paper on Various Search Engines (Google, Yahoo, Altavista, Ask and Bing). *International Journal on Recent and Innovation Trends in Computing and Communication*, 4 (8),190- 195.

Fatima, A., Luca, C., & Wilson, G. (2014). New Framework for Semantic Search Engine**.** *International Conference on Computer Modelling and Simulation*. UKSim-AMSS,445-450. DOI:10.1109/UKSim.2014.114

Gayar, M.M.E., Mekky N., & Atwan, A. (2015). Efficient Proposed Framework for Semantic Search Engine using New Semantic Ranking Algorithm. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 6(8),136- 143.

Goutam, R. K., & Dwivedi, S. K. (2012). Performance Evaluation of search engines via user efforts Measures. *International Journal of Computer Science Issues*, 9(2),432- 437.

Guha, R., McCool, R. ,& Miller, E. (2003). *Semantic Search*. Retrieved 23 October, 2017 from: https://dl.acm.org/citation.cfm?id=775250

Signorini, A., & Imielinski, T. (2009). If You Ask Nicely, I will Answer: Semantic Search and Today's Search Engines. 2009 IEEE International Conference on Semantic Computing, 184-191.

Kousha, K. (2001). *Internet Tools: Principles Skills and Facilities for Searching on World Wide Web.* Tehran: ketabdar.

Kumar, D., & Mishra, R. (2015). Deep Web Performance Enhance on Search Engine. 2015 International Conference on Soft Computing Techniques and Implementations- (ICSCTI), Hybrid Intelligence. 137- 140.

Lancaster, F. W. (2002). indexing and abstracting(A.Gilvari Trans.). Tehran: chapar.

Mala, V., & Lobiyal, D. K. (2016). Semantic and keyword based web techniques in information retrieval. 2016 International Conference on Computing, Communication and Automation (ICCCA). https://doi.org/10.1109/ccaa.2016.7813724

Malve, A., & Chawan, P. M. (2015). A Comparative Study of Keyword and Semantic based Search Engine. *International Journal of Innovative Research in Science,Engineering and Technology*, 11(4),11156-11161.

Nadi Ravandi, S., & Hariri, N. (2016.). *Information Retrieval Systems* .Tehran: Ketabdar.

Nandy, D., Gohel, H., Kalyani, F. ,& Khakhariya, M. (2016). DuckDuckGo vs. Google search engine comparative analysis. *International Journal of Science Research and Technology*, 2(1), 37-41.

Negi, Y. S., & Kumar, S. (2014). A Comparative Analysis of Keyword-and Semantic-Based Search Engines. In *Intelligent Computing, Networking, and Informatics*(pp.727-736).

Journal of Knowledge-Research Studies (JKRS)

Vol 1

Issue 2

Spring 2022

Nowruzi, M., & Taherian, M. (2012). Semantic web. Tehran: fors.

Ramachandran, A., & Sujatha, R. (2011). Semantic search engine: A survey. *Int. J. Comp. Tech. Appl*, 2 (6),1806-1811.

Rashid, J., & Wasif Nisar, M. (2016). A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines. *I.J. Information Technology and Computer Science,* 10(10),82- 89.

Rathee, D., Atri, J., Rathee, J., & Nath, R. (2013). A Framework to Evaluate Search Engines. *IJCSC*, 4 (2), 184-190.

Sahu, S. K., Mahapatra,and, D.P., & Balabantaray, R.C. (2015).Comparative Study of Different Search Engines in Context of Features and Semantics, *International Journal of All Research Education and Scientific Methods*, 14(1), 15-26.

Sayed, A., & Al Muqrishi, A. (2017). IBRI-CASONTO: Ontology-based semantic search engine, *Egyptian Informatics Journal*,18 (3), 181- 192.

Sheela, A.C., & Jayakumar, C.T. (2019). Comparative Study of Syntactic Search Engine and Semantic Search Engine: A Survey. 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), 1, 1-4.

Shiri, A. A. (1994). Evaluate recall and precision of information retrieval systems. *Research on Information Science and Public Libraries, 4 (1,2), 38-45. [In Persian].*

Singh, J., & Sharan, A. (2013). A Comparative Study between Keyword and Semantic Based Search Engines. *International Conference on Cloud, Big Data and Trust 2013*: 130- 134.

Solomou, G. & Koutsomitropoulos, D. (2015). Towards an evaluation of semantic searching in digital repositories: a DSpace case-study. *Electronic library and information systems*, 49 (1), 63-90.

Sudeepthi, G., Anuradha, G., Surendra, M.,& Babu, P. (2012). A Survey on Semantic Web Search Engine. *IJCSI International Journal of Computer Science Issues*, 9(2), 241-245.

Tümer, D., Shah, M.A., & Bitirim, Y. (2009). An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia. 2009 Fourth International Conference on Internet Monitoring and Protection, 51-55.

Yadav, A., & Maheshwari, S. (2017). Building of Intellectual Web Search Engines through Semantic Web. *International Journal of Latest Trends in Engineering and Technology*, 7 (3),384-391.

Journal of Knowledge-Research Studies

Journal of Knowledge-Research Studies (JKRS)

Vol 1

Issue 2

Spring 2022

نشریه مطالعات دانش پژوهی

بهاری ورزنه، حسین (۱۴۰۱). مقایسه عملکرد بازیابی اطلاعات موتورهای جستجوی معنایی و کلیدواژه‌ای بر اساس جستجوی عبارتی. *نشریه مطالعات دانش‌پژوهی*، ۱(۲): ۱۰۲-۱۱۷.

DOI: 10.22034/JKRS.2022.51351.1014
URL: https://jkrs.tabrizu.ac.ir/article_15085.html

این مقاله به‌صورت دسترسی باز و با لایسنس CC BY NC کریتیوکامانز قابل‌استفاده است.

مقایسه عملکرد بازیابی اطلاعات موتورهای جستجوی معنایی و کلیدواژه‌ای بر اساس جستجوی عبارتی

حسین بهاری ورزنه [۱]

۱. کارشناس ارشد علم اطلاعات و دانش‌شناسی دانشگاه اصفهان، و مسئول کتابخانه تخصصی ورزش و سلامت ابرار شهرداری اصفهان(نویسنده مسئول)
bahari_2062@yahoo.com

تاریخ دریافت ۱ اردیبهشت ۱۴۰۱ **تاریخ پذیرش: ۲۷ خرداد ۱۴۰۱**

چکیده

هدف: هدف از انجام این پژوهش مقایسه عملکرد بازیابی اطلاعات موتورهای جستجوی معنایی و کلیدواژه بر اساس جستجوی عبارت (ساده و پیچیده) است.

روش‌شناسی: جامعه پژوهش کاربردی و نیمه تجربی حاضر شامل تمام موتورهای جستجوی فعال در وب است. نمونه‌های پژوهش بر اساس نمونه‌گیری طبقه‌ای تصادفی و نمونه‌گیری هدفمند انتخاب شدند. ابزار گردآوری داده‌های دو چک‌لیست محقق ساخته شامل ۱۰ پرس و جوی عبارتی ساده و پیچیده است.

یافته‌ها: بینگ و کلاز (با میزان دقت مشابه ۵۳٪)، داک‌داک‌گو و یاهو به ترتیب بیشترین دقت را در جستجوی عبارات ساده داشتند. در جستجوی عبارات پیچیده نیز به ترتیب بینگ، داک‌داک‌گو، یاهو و کلاز بیشترین میزان دقت را داشتند. به‌طورکلی، بینگ، داک‌داک‌گو، کلاز و یاهو به ترتیب بیشترین میزان دقت را دارند. همچنین، میزان میانگین کل دقت موتورهای جستجوی کلیدواژه‌ای از موتورهای جستجوی معنایی بیشتر است.

نتایج: موتور جستجوی کلیدواژه‌ای بینگ بهترین عملکرد را نسبت به سه موتور جستجوی معنایی و کلیدواژه‌ای دیگر دارد. موتورهای جستجوی معنایی بااینکه مدعی هستند توانایی‌های بیشتری در بازیابی اطلاعات مرتبط نسبت به موتورهای جستجوی کلیدواژه‌ای دارند اما در این بررسی مشخص شد کلاز و داک‌داک‌گو در جستجوی عبارتی نسبت به موتورهای جستجوی کلیدواژه‌ای برتری ندارند. این ابزارها به‌عنوان موتورهای جستجوی وب معنایی عملکرد قابل قبولی از خود نشان ندادند و به نظر می‌رسد باید مسیر طولانی را طی نمایند تا به موتورهای جستجوی معنایی واقعی تبدیل شوند و برای نیل به این مهم لازم است از امکانات، ابزارها، ماژول‌ها، و فناوری‌های نوظهور عصر جدید از قبیل یادگیری ماشین، یادگیری عمیق، تلفیق این ماژول‌ها با تکنیک‌های متن‌کاوی، داده‌کاوی، و... بهره گیرند.

اصالت: تاکنون در موتورهای جستجوی معنایی و کلیدواژه‌ای نمونه، عملکرد جستجوی عبارتی مقایسه نشده است. و در این راستا، پژوهشگر تلاش کرده است تا با یک بررسی دقیق به یک نتیجه واقعی دست یابد تا موتور جستجوی قدرتمند در این نوع جستجو را به کاربران معرفی کند.

کلیدواژه‌ها: ***بازیابی اطلاعات، جستجوی عبارتی، عملکرد، موتورهای جستجوی کلیدواژه‌ای، موتورهای جستجوی معنایی***

۱. مقدمه

اینترنت را می‌توان یکی از بزرگ‌ترین دستاوردهای بشر دانست که امروزه تمام جوانب زندگی انسان را تحت تأثیر قرار داده است (نوروزی و طاهری، ۱۳۹۰: ۱۲). رشد منابع اطلاعاتی سبب شده است تا اهمیت بازیابی و دسترسی به اطلاعات بیشتر از قبل شود. ظهور وب و شخصی‌سازی تولید اطلاعات یعنی فراگیر شدن تولید اطلاعات در میان عموم مردم باعث شده است تا حوزه بازیابی اطلاعات به یک موضوع مهم تبدیل شود (نادی راوندی و حریری، ۱۳۹۴: ۱۷-۱۸). برای بازیابی اطلاعات از وب، موتورهای جستجو ضروری هستند (گوتام و دیویدی[1]، ۲۰۱۲). موتور جستجو درواقع برنامه‌ای است که اجزای یک پایگاه داده را مورد جستجو و شناسایی قرار می‌دهد و با کلمات و نویسه‌هایی که مدنظر کاربر است مطابقت می‌دهد. موتورهای جستجو از طریق اچ.تی.ام.ال.[2] به ذخیره اطلاعات درباره صفحات وب بازیابی شده می‌پردازند. این صفحات توسط خزنده‌های وب که لینک‌ها را روی سایت دنبال می‌کنند بازیابی می‌شوند. سپس محتویات هر صفحه را تجزیه‌وتحلیل می‌کنند تا مشخص شود چگونه باید آن را نمایش دهند (به‌عنوان مثال، می‌توانند کلمات را از عناوین، محتوای صفحه یا از فیلدهای خاصی که ابر برچسب دارند استخراج کنند). اطلاعات مربوط به صفحات وب در پایگاه داده فهرست و برای استفاده برای پرس‌وجوهای بعدی ذخیره می‌شود (کومار و میشرا[3]، ۲۰۱۵). به‌طور معمول، موتورهای جستجو هزاران صفحه را در پاسخ به یک پرس‌وجو ارائه می‌دهند ولی عملاً امکان دسترسی به تمام این اسناد وجود ندارد (گوتام و دیویدی، ۲۰۱۲). وب کنونی پایگاه داده جهانی است که وجود محدودیت‌هایی نظیر ا. محتوای فاقد ساختار مناسب برای نمایش اطلاعات، ۲. ابهام در اطلاعات به خاطر پیوند ضعیف بین آن‌ها، ۳. عدم توانایی درک اطلاعات ارائه‌شده توسط ماشین‌ها به دلیل عدم وجود فرمت جهانی، و ۴. فقدان انتقال خودکار اطلاعات در آن باعث بازیابی رکوردهای نامرتبط زیادی می‌شود (سودفی، آنوردها و پرساد باب[4] و، ۲۰۱۲). علاوه بر آن، وجود میلیاردها پایگاه داده مقدار اطلاعات را در هرسال افزایش می‌دهد. امروزه موتورهای جستجوی بسیاری وجود دارد اما چالش اصلی این موتورهای جستجو این است که اکثر آن‌ها نمی‌توانند به‌صورت هوشمندانه اطلاعات معنایی را بازیابی کنند (گایار، مکی و آتوان[5]، ۲۰۱۵). اطلاعات جستجو شده توسط این موتورهای جستجو با بازیابی بالا و دقت کم یا بدون بازیابی (زمانی که هیچ نتیجه مرتبطی وجود نداشته باشد) همراه است (فاطیما، لوکا و ویلسون[6]، ۲۰۱۴). فنّاوری وب معنایی یک راه‌حل برای این مشکل است که داده‌ها را در یک فرمت قابل‌خواندن نگه می‌دارد و به ماشین‌ها کمک می‌کند تا این اطلاعات را بر اساس معانی با پرس و جوی کاربر منطبق

1 . Goutam & Dwivedi
2 . hypertext markup language (HTML)
3 . Kumar & Mishra
4 . Sudeepthi, Anuradha & Surendra Prasad Babu
5 . gayar, Mekky & Atwan
6 . Fatima, Luca & Wilson

سازند (گایار، مکی و آتوان، ۲۰۱۵). وب معنایی گسترش وب فعلی است که در آن معنای اطلاعات به‌خوبی تعریف شده است (گوها، مک کول و میلر[۱]، ۲۰۰۳). تیم برنرزلی[۲] وب معنایی را این‌گونه تعریف می‌کند: «وب معنایی یک فرایند رایج را فراهم می‌کند که اجازه می‌دهد داده‌ها در میان برنامه‌های کاربردی، شرکت‌ها و محدوده‌های اجتماعی به اشتراک گذاشته شوند و دوباره و دوباره استفاده شوند» (چودهاری[۳]، ۲۰۱۴). دستاورد وب معنایی، موتورهای جستجوی معنایی هستند. موتورهای جستجوی معنایی به‌طور مؤثر داده‌ها را جستجو و اطلاعات معناداری را به‌طور هوشمندانه بازیابی می‌کنند (یاداو و مهشواری[۴]، ۲۰۱۷). موتورهای جستجوی معنایی دارای مزایایی نسبت به موتورهای جستجو در وب از دید کاربران هستند. امروزه بشر نیاز به یافتن سریع پاسخ مرتبط با پرس و جوی خود دارد و از این حیث موتورهای جستجو معنایی مفید خواهند بود (راماچاندران و سوجاتا[۵]، ۲۰۱۱). موتورهای جستجوی معنایی برخلاف موتورهای جستجوی سنتی که از الگوریتم رتبه‌بندی صفحات استفاده می‌کنند تا رتبه‌بندی را به لینک خاصی اختصاص دهند و نتایج مرتبطی ارائه دهند، از هستی‌شناسی استفاده می‌کنند. به‌طوری که نتایج معنی‌دار و دقیقی در زمان کمتری بازیابی می‌شود (ملوه و چاوان[۶]، ۲۰۱۵). آن‌ها بر اساس رویکرد معنایی کار می‌کنند (ساهو، ماهاپاترا و بالابانتاری[۷]، ۲۰۱۵) و خواسته کاربر را تجزیه‌وتحلیل کرده و با استفاده از استدلال منطقی نتایج دقیق‌تری را بازیابی می‌کنند. موتورهای جستجوی معنایی، معنای واقعی و نهفته در پرسش کاربر را تشخیص داده و با توجه به آن، نتایج را ارائه می‌دهند (دری، ۱۳۹۳). بنا براین، همان‌طور که گفته شد اطلاعات معنی‌دار و دقیقی را به کاربران ارائه می‌دهند (رشید و وصیف نیثار[۸]، ۲۰۱۶). این موتورهای جستجو با استفاده از جستجوی معنایی به بازیابی اطلاعات می‌پردازند. جستجوی معنایی یک برنامه از وب معنایی برای جستجو است که تلاش می‌کند با استفاده از داده‌های وب معنایی، نتایج جستجوی سنتی را بهبود بخشد (گوها، مک کول و میلر، ۲۰۰۳). جستجوی معنایی مجموعه‌ای جدید از پرس‌وجوها را بر اساس قدرت موتورهای استنتاج فراهم می‌کند درحالی‌که در جستجوی کلیدواژه‌ای این امکان وجود ندارد (سلمو و کتسومیتروپلس[۹]، ۲۰۱۵). این نوع جستجو دقت و صحت جستجوی اطلاعات مربوط به پرس و جوی کاربر را در سطح بالایی بهبود می‌بخشد و محتوای دقیقی را بر اساس آنچه کاربر می‌خواهد درباره آن بداند فراهم می‌کند (سودفی، انوردا و سرندرا پرساد بابو، ۲۰۱۲).

1 . Guha, McCool & Miller
2 . Tim Berners-Lee
3 . Choudhury
4 . Yadav & Maheshwari
5 . Ramachandran, A., & Sujatha
6 . Malve & Chawan
7 . Sahu, Mahapatra & Balabantaray
8 . Rashid & Wasif Nisar,
9 . Solomou & Koutsomitropoulos

برای ارزیابی عملکرد ابزارهای کاوش در بازیابی اطلاعات از دنیای وب معیارهایی توسعه یافته و مورد استفاده قرار گرفته است. ربط[1]، یکی از این معیارهاست که برای ارزیابی عملکرد نظام‌های بازیابی اطلاعات و کارایی کاوش‌های انجام‌شده به کار می‌رود و به‌وسیله دو مقیاس جامعیت و مانعیت سنجیده می‌شود. ارزیابی کارایی یک نظام بازیابی اطلاعات صرف‌نظر از اینکه با چه نوع پایگاه اطلاعاتی مواجه باشیم، بدون بررسی دقت و بازیافت ناتمام خواهد ماند و به‌درستی نمی‌توان درباره کارایی آن قضاوت کرد (کوشا، ۱۳۸۱، ۸۰ و ۸۸). دقت با مدارک مربوط بازیابی شده ارتباط مستقیم دارد و نشان می‌دهد چه اندازه نظام در بازیابی اطلاعات مرتبط موفق عمل کرده است؛ یعنی در میان مدارک بازیابی شده چه تعداد مدرک به نیاز اطلاعاتی کاربر مربوط است (نادی راوندی و حریری، ۱۳۹۴، ۱۵۴). دقت یا مانعیت نسبت تعداد مدارک مربوط بازیابی شده به تعداد کل مدارک بازیابی شده است و بازیافت یا جامعیت نسبت تعداد مدارک مربوط بازیابی شده به تعداد کل مدارک مربوط موجود در مجموعه است (شیری، ۱۳۷۳).

ازآنجایی‌که کاربران اینترنت اطلاع چندانی از عملکرد موتورهای جستجوی معنایی و کلیدواژه‌ای ندارند و با توجه به اینکه معلوم نیست این موتورهای جستجو در بازیابی رکوردهای مربوط به پرس‌وجوهای عبارتی (ساده و پیچیده) چه عملکردی دارند، لزوم اجرای پژوهش‌هایی جهت بررسی عملکرد آن‌ها در این خصوص می‌تواند در انتخاب موتورهای جستجوی موفق جهت بازیابی اطلاعات مرتبط با عبارات مورد جستجوی کاربران کمک‌کننده باشد. لذا پژوهش حاضر قصد دارد در راستای دستیابی به این مهم و پاسخگویی به سؤالات زیر به بررسی و مقایسه عملکرد جستجوی عبارات (ساده و پیچیده) در این موتورهای جستجو بپردازد:

۱. میزان دقت رکوردهای بازیابی شده بر اساس پرس و جوی عبارات ساده در موتورهای جستجوی موردبررسی چقدر است؟

۲. میزان دقت رکوردهای بازیابی شده بر اساس پرس و جوی عبارات پیچیده در موتورهای جستجوی موردبررسی چقدر است؟

۳. کدام دسته از موتورهای جستجوی موردبررسی میزان دقت بیشتری دارند؟

1. relevance

۲. پیشینه پژوهش

در بیشتر پژوهش‌ها عملکرد بازیابی اطلاعات موتورهای جستجو با استفاده از مقیاس‌های بازیافت و دقت سنجیده می‌شود. مرور نوشتارها نشان می‌دهد که پژوهش‌های کمی همانند پژوهش تومر، شاه و بیتیریم[۱] در سال ۲۰۰۹ انجام شد. آن‌ها یک ارزیابی تجربی از عملکرد موتورهای جستجوی معنایی و کلیدواژه‌ای به عمل آوردند. بعد از آن‌ها، آنداگو، فوبی و تانون[۲] در سال ۲۰۱۰ نتایجی در این حوزه موضوعی منتشر کردند. الداباغ[۳] (۲۰۱۲) نیز پژوهشی مشابه پژوهش آنداگو و همکارانش انجام داد و عملکرد موتورهای جستجوی معنایی و کلیدواژه‌ای را باهم مقایسه کرد. سینگ و شاران[۴] نیز در سال ۲۰۱۳ یک مطالعه تطبیقی بین موتورهای جستجوی معنایی و کلیدواژه‌ای انجام دادند. احمدخان، سنگروا و رحمان[۵] (۲۰۱۴) نتایج یک ارزیابی از عملکرد موتورهای جستجوی معنایی و کلیدواژه‌ای را گزارش دادند. در همان سال نگی و کومار[۶] یک تجزیه‌وتحلیل تطبیقی بین موتورهای جستجوی معنایی و کلیدواژه‌ای انجام دادند. مالا و لوبیال[۷] (۲۰۱۶) نیز تکنیک‌های وب معنایی و کلیدواژه‌ای را در بازیابی اطلاعات بررسی کردند. سید و الموکریشی[۸] در سال ۲۰۱۷ یک موتور جستجوی مبتنی بر هستی‌شناسی را بررسی و عملکرد آن را با سه موتور جستجوی معنایی و کلیدواژه‌ای مورد مقایسه قراردادند. یک سال بعد عزیزان و همکاران[۹] (۲۰۱۸)، عملکرد بازیابی اطلاعات چهار موتور جستجوی گوگل، بینگ، یاهو، و داک داک گو را برای بازیابی اطلاعات مربوط به هشت پرس‌وجو را موردبررسی و مقایسه قراردادند. شلا و جایاکومار[۱۰] (۲۰۱۹) به مطالعه تطبیقی عملکرد بازیابی اطلاعات پنج موتور جستجوی داک داک گو، بینگ، گوگل، یاهو، هاکیا، و ویکی‌پدیا پرداختند. جاتوانی، تومر، و جانگرا[۱۱] (۲۰۲۰) به ارزیابی و مقایسه عملکرد بازیابی اطلاعات موتورهای جستجوی گوگل، بینگ، و داک داک گو از طریق دسته‌بندی‌های مختلف پرس‌وجو پرداختند. علیو و یحیی[۱۲] (۲۰۲۱) به بررسی دقت موتورهای جستجوی مبتنی بر گراف دانش: گراف دانش گوگل، بینگ و والفرام آلفا پرداختند) به بررسی عملکرد موتورهای جستجوی معنایی و کلیدواژه‌ای از این طریق پرداخته‌اند. در جدول ۱، خلاصه‌ای از این پژوهش‌ها ارائه شده است.

1 . Tumer, Shah & Bitirim
2 . Andago Phoebe & Thanoun
3. Aldabbagh
4 . Singh & Sharan
5 . Ahmad Khan, Sangroha & Rahman
6 . Negi & Kumar
7 . Mala & Lobiyal
8 . Sayed & Al Muqrishi
9 . Azizan et al
10 . Sheela & Jayakumar
11 . Jatwani, Tomar, & Dhingra
12 . Aliyu & Yahaya

جدول ۱. خلاصه پیشینه پژوهش

ردیف	پژوهشگر/ تاریخ	روش‌شناسی	نتیجه
۱	تومر، شاه و بیتیریم (۲۰۰۹)	ده پرسش از موضوعات مختلف و چهار عبارت، با نحو متفاوت ولی با معانی مشابه، برای پرس جو تعیین شدند. بیست نتیجه اول در هر پرس‌وجو به‌صورت «مرتبط» یا «نامرتبط» ارزیابی شدند.	یاهو، هاکیا، گوگل و ام اس ان به ترتیب بهترین عملکرد را داشتند.
۲	آنداگو، فوبی و تانون (۲۰۱۰)	ده پرسش از سی دانشجو گردآوری شد و بیست نتیجه اول مطابق با نظر آن‌ها بر اساس «مرتبط» یا «نامرتبط» بودن مورد ارزیابی قرار گرفت.	گوگل عملکرد بهتری نسبت به هاکیا از خود نشان داد.
۳	الداباغ (۲۰۱۲)	ده پرسش از موضوعات مختلف انتخاب و بیست نتیجه اول در هر پرس‌وجو بررسی شدند. جمعاً چهار هزار اسناد بازیابی شده به‌عنوان رکوردهای«مرتبط»، «غیر مرتبط»، «تکراری» و «غیرقابلقابل‌دسترس» طبقه‌بندی شدند.	هاکیا، گوگل، یاهو و اماس‌ان به ترتیب بهترین عملکرد را داشتند.
۴	سینگ و شاران (۲۰۱۳)	ده پرسش که حاوی موضوعات مختلف که شامل یک یا دو اصطلاح هستند به‌طور تصادفی مشخص شدند. بیست نتیجه اول بازیابی شده د هر پرس‌وجو بررسی و به‌صورت«مرتبط» یا «نامرتبط» ارزیابی شدند.	به ترتیب بینگ، داک‌داک‌گو، گوگل، یاهو و هاکیا بهترین عملکرد را داشتند.
۵	احمدخان، سنگروا و رحمان (۲۰۱۴)	ده پرسش از موضوعات مختلف انتخاب و ده نتیجه اول بازیابی شده در هر پرس‌وجو به‌صورت «مرتبط» یا «نامرتبط» مورد ارزیابی قرار گرفتند.	بینگ، داک‌داک‌گو، گوگل، یاهو و لکسه[1] به ترتیب قوی‌تر هستند.
۶	نگی و کومار (۲۰۱۴)	ده پرسش از موضوعات مختلف انتخاب شدند و بیست نتیجه اول بازیابی شده در هر پرس‌وجو به‌صورت «مرتبط» یا «نامرتبط» ارزیابی شدند.	بینگ، داک‌داک‌گو، گوگل، یاهو و هاکیا به ترتیب بهترین عملکرد را داشتند.
۷	مالا و لوبیال (۲۰۱۶)	ده پرسش از موضوعات مختلف شامل یک یا دو اصطلاح به‌صورت تصادفی انتخاب شدند. بیست نتیجه اول بازیابی شده به‌صورت «مرتبط» یا «نامرتبط» ارزیابی شدند.	به ترتیب داک‌داک‌گو، گوگل، یاهو، بینک، ویکی‌پدیا و هاکیا بهترین عملکرد را داشتند.

1 . Lexxe

۸	سید و الموکریشی (۲۰۱۷)	چهل پرسش از موضوعات مختلف انتخاب شدند و چهل نتیجه اول بازیابی شده از هر پرس‌وجو به‌صورت «مرتبط» یا «غیر مرتبط» مورد ارزیابی قرار گرفتند.	موتورهای جستجوی IBRICASONTO، گوگل، والفرام آلفا[1] و کانجین[2] به ترتیب قوی تر بودند.
۹	عزیزان و همکاران (۲۰۱۸)	هشت پرس‌وجو از موضوعات مختلف انتخاب شدند و بیست نتیجه اول بازیابی شده در هر پرس‌وجو به‌صورت «مرتبط» یا «نامرتبط» ارزیابی شدند.	گوگل از سه موتور جستجوی دیگر بهتر عمل کرده است. پس‌ازآن به ترتیب یاهو، داک داک گو، و بینگ قرار گرفتند.
۱۰	شلا و جایاکومار (۲۰۱۹)	برای انجام این پژوهش از ده پرس‌وجو از موضوعات مختلف استفاده شد. بیست نتیجه اول بازیابی شده از هر پرس‌وجو به‌صورت «مرتبط» یا «غیر مرتبط» مورد ارزیابی قرار گرفتند.	به ترتیب ویکی‌پدیا، داک داک گو، گوگل، یاهو، بینگ، و هاکیا بهترین عملکرد را از خود نشان دادند.
۱۱	علیو و یحیی[3] (۲۰۲۱)	۱۴۹ پرس‌وجو از موضوعات مختلف مورد جستجو قرار گرفت و گراف دانش ارائه‌شده توسط سه موتور جستجوی نمونه بر اساس منطق کلاسیک بررسی شد.	گراف دانش گوگل عملکرد بهتری را در این خصوص از خود نشان داده است. بعدازآن به ترتیب بینگ و وافرام آلفا قرار گرفتند.

نشریه مطالعات دانش‌پژوهی

صفحه۱۰۶ |

دوره ۱، شماره ۲

بهار ۱۴۰۱

پیاپی ۲

جدول ۱ گویای این است که در اکثر این پژوهش‌ها ده نمونه پرس‌وجو برای بررسی عملکرد موتورهای جستجوی موردبررسی انتخاب شد و ارتباط بیست نتیجه اول بازیابی شده از هر پرس‌وجو موردبررسی قرار گرفت. همچنین، همه این پژوهش‌ها به‌صورت صفر و یک یا کلاسیک به ارزیابی مرتبط یا نامرتبط بودن نتایج بازیابی شده در هر موتور جستجو پرداختند. به‌طورکلی، مرور نوشتارها نشان داد پژوهشی به‌طور انحصاری «به بررسی و مقایسه عملکرد جستجوی عبارتی در موتورهای جستجوی کلیدواژه‌ای و موتورهای جستجوی معنایی» نپرداخته است. بر این اساس، لزوم انجام پژوهش حاضر دوچندان می‌شود.

۳. روش‌شناسی پژوهش

پژوهش کاربردی حاضر ازآن‌جهت که پژوهشگر توانایی کنترل تمام متغیرهای مؤثر بر نتایج بازیابی شده را ندارد در دسته پژوهش‌های نیمه تجربی قرار می‌گیرد. جامعه پژوهش شامل تمام موتورهای جستجوی فعال در وب است. پژوهشگر ابتدا بر اساس نمونه‌گیری طبقه‌ای تصادفی

1 . walffrom alpha
2 . kngine
3 . Aliyu & Yahaya

موتورهای جستجو را به دودسته موتورهای جستجوی معنایی و کلیدواژه‌ای تقسیم کرد. سپس در این دودسته موتور جستجو بر اساس برتری و کاربرد موتورهای جستجو در آثار پژوهشی پیشین و میزان محبوبیت آن‌ها نزد کاربران وب جهان‌گستر دست به نمونه‌گیری هدفمند زد و از میان موتورهای جستجوی معنایی، دو موتور جستجوی داک‌داک‌گو و کلاز و از میان موتورهای جستجوی کلیدواژه‌ای دو موتور جستجوی یاهو و بینگ را به‌عنوان نمونه‌های موردبررسی انتخاب کرد. لازم به ذکر است که معنایی بودن و کلیدواژه‌ای بودن موتورهای جستجوی نام‌برده با استناد به آثار پژوهشی معتبر پیشین انجام گرفته است. برای گردآوری ادبیات و پیشینه پژوهش نیز از روش کتابخانه‌ای استفاده شد. ابزار گردآوری داده‌ها دو چک‌لیست محقق ساخته است که طبق جدول ۲، هرکدام شامل ۵ نمونه پرس‌وجو برای جستجوی عبارتی ساده و پیچیده می‌باشد. عبارات ساده دارای حرف‌اضافه نیستند و در تیپ ۱ قرار دارند؛ و عبارات پیچیده دارای حرف‌اضافه هستند و در تیپ ۲ قرار دارند. این عبارات بر اساس سرعنوان‌های موضوعی فارسی و در حوزه‌های موضوعی روانشناسی، سلامت، پزشکی، علوم پایه، فناوری، مذهب، و سیاست برگزیده شدند. و روایی آنا توسط دو متخصص موردبررسی قرار گرفت.

جدول ۲. تیپ بندی نمونه‌های پرس‌وجو

نمونه‌های پرس‌وجو			
ردیف	عبارات ساده (تیپ ۱)	ردیف	عبارات پیچیده (تیپ ۲)
۱	Sigmund Freud's biography	۱	Symptoms of heart attack
۲	Pear Benefits	۲	Year of Jesus born
۳	Water Elements	۳	Traditional of Japanese in New Year
۴	Penicillin Discoverer	۴	Current president of United States
۵	Semantic retrieval	۵	Number of killed in Vietnam War

در چهار موتور جستجوی نمونه، میزان ارتباط ۲۰ نتیجه اول بازیابی شده در هر پرس‌وجو موردبررسی قرار گرفت و با استفاده از فرمول دقت ارائه‌شده توسط لنکستر (۱۳۸۲)، مانعیت این موتورهای جستجو بر اساس عبارات تیپ ۱ و ۲ محاسبه شد. برای هر موتور جستجو یک میانگین کل دقت به دست آمد که بر اساس آن عملکرد آن‌ها مورد مقایسه قرار گرفت. از میزان دقت کل به دست آمده برای موتورهای جستجوی هر دودسته نیز میانگین گرفته شد تا از این طریق عملکرد دودسته موتور جستجو باهم مورد مقایسه قرار گیرد. برای تجزیه‌وتحلیل داده‌ها از آمار توصیفی، و برای رسم نمودارها از نرم‌افزار اکسل استفاده شده است.

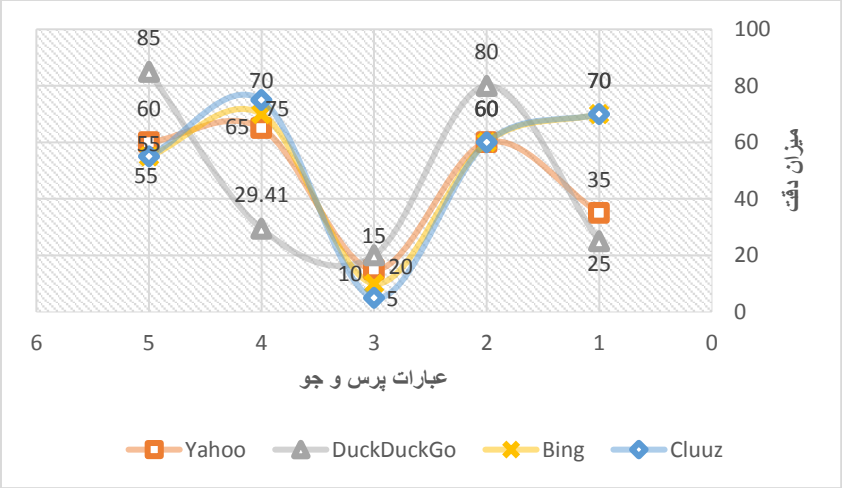
$$\text{میزان دقت} = \frac{\text{تعداد مدارک بازیابی شده مرتبط}}{\text{تعداد کل مدارک بازیابی شده}} \times 100$$

۴. یافته‌ها

در این بخش از پژوهش یافته‌های حاصل‌شده جهت پاسخگویی به سؤالات پژوهش ارائه شده‌اند. ابتدا میزان ارتباط رکوردهای بازیابی شده حاصل از پرس و جوی دو تیپ عبارت در نظر گرفته در چهار موتور جستجوی معنایی و کلیدواژه‌ای نمونه، توسط پژوهشگر موردبررسی قرار گرفت. در ادامه میانگین کل این داده‌ها برای هر موتور محاسبه شد و پس‌ازآن میانگین کل دقت برای هر دسته‌موتور به دست آمد و طبق آن عملکرد بازیابی اطلاعات در این موتورهای جستجو مورد مقایسه قرار گرفت. این یافته‌ها در ادامه ارائه شده‌اند.

الف. میزان دقت رکوردهای بازیابی شده بر اساس پرس و جوی عبارات ساده در موتورهای جستجوی موردبررسی چقدر است؟

برای پاسخگویی به این سؤال، میزان دقت حاصل از بررسی ۲۰ نتیجه اول بازیابی شده بر اساس عبارت‌های ساده در چهار موتور جستجوی نمونه طبق نمودار ۱، محاسبه شد و موردبررسی قرار گرفت.



نمودار ۱. میزان دقت موتورهای جستجو بر اساس پرس‌وجوهای تیپ ۱ (عبارات ساده)

نمودار ۱ یک خلاصه گرافیکی از میزان دقت حاصل از بررسی ۲۰ نتیجه اول از پرس‌وجوهای تیپ ۱ را نشان می‌دهد. همان‌طور که مشخص است، حداقل دقت کمتر از ۱۰٪ بوده و مربوط به پرس و جوی ۳ است و در موتور جستجوی کلاز دیده می‌شود. حداکثر دقت نیز نزدیک به ۹۰٪ بوده و مربوط به پرس و جوی ۵ است و در موتور جستجوی داک داک گو دیده می‌شود. در پرس و جوی ۳ نیز چهار موتور جستجو میزان دقت کمتر از ۲۰٪ را به دست آوردند. در پرس و جوی ۱، دو موتور جستجوی کلاز و بینگ میزان دقت مشابهی از خود نشان دادند، در پرس و جوی ۲ نیز سه موتور جستجوی کلاز، بینگ و یاهو میزان دقت مشابهی به دست آوردند و در پرس و جوی ۵، دو موتور جستجوی کلاز و بینگ از میزان دقت مشابه برخوردارند.

نشریه مطالعات دانش‌پژوهی

صفحه ۱۰۸ |

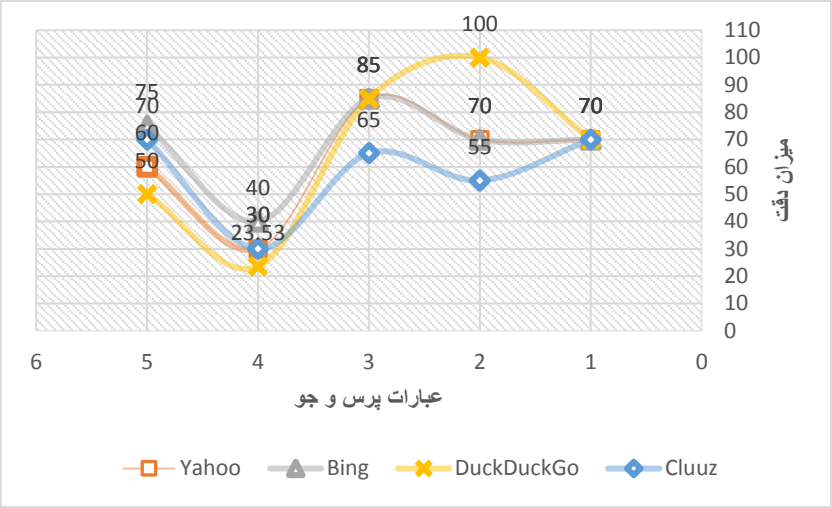
دوره ۱، شماره ۲

بهار ۱۴۰۱

پیاپی ۲

ب. میزان دقت رکوردهای بازیابی شده بر اساس پرس و جوی عبارات پیچیده در موتورهای جستجوی موردبررسی چقدر است؟

برای پاسخگویی به سؤال دوم پژوهش، میزان دقت حاصل از بررسی ۲۰ نتیجه اول بازیابی شده بر اساس عبارت‌های پیچیده در چهار موتور جستجوی نمونه طبق نمودار ۲، محاسبه شد و موردبررسی قرار گرفت.



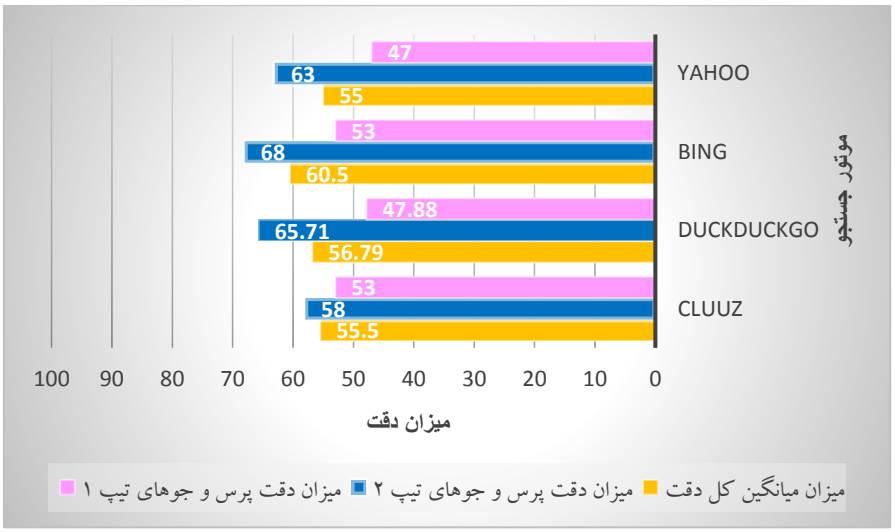
نمودار ۲. میزان دقت موتورهای جستجو بر اساس پرس‌وجوهای تیپ ۲ (عبارات پیچیده)

نمودار ۲ یک خلاصه گرافیکی از میزان دقت حاصل از بررسی ۲۰ نتیجه اول از پرس‌وجوهای تیپ ۲ را نشان می‌دهد. همان‌طور که مشخص است، حداقل دقت نزدیک به ۲۰٪ بوده و مربوط به پرس و جوی ۴ است و در موتور جستجوی داک داک گو دیده می‌شود. حداکثر دقت نیز ۱۰۰٪ بوده و مربوط به پرس و جوی ۲ است و در موتور جستجوی داک داک گو دیده می‌شود. در پرس و جوی ۴ نیز چهار موتور جستجو کمترین میزان دقت را به دست آوردند. در پرس و جوی ۱ چهار موتور جستجو میزان دقت مشابه به دست آوردند، در پرس و جوی ۲، دو موتور جستجوی بینگ و یاهو میزان دقت مشابه دارند و در پرس و جوی ۳ نیز سه موتور جستجوی داک داک گو، بینگ و کلاز میزان دقت مشابه از خود نشان دادند.

ج. کدام دسته از موتورهای جستجوی موردبررسی میزان دقت بیشتری دارند؟

برای پاسخگویی به سومین سؤال پژوهش، طبق نمودار ۳، میزان میانگین دقت حاصل از بررسی ۲۰ نتیجه اول بازیابی شده بر اساس عبارت‌های ساده و پیچیده، به‌طور جداگانه در چهار موتور جستجوی نمونه محاسبه شد. و میانگین کل دقت هر موتور جستجو برای مقایسه عملکرد آن‌ها محاسبه شد.

.



نمودار ۳. میزان میانگین دقت موتورهای جستجو

همان‌گونه که در نمودار ۳ مشهود است موتور جستجوی بینگ و کلاز با میزان دقت مشابه ۵۳٪ بیشترین میزان دقت را در جستجوی عبارات تیپ ۱ که عبارات ساده محسوب می‌شوند، به دست آوردند. بعد از آن‌ها، موتور جستجوی داک‌داک‌گو قرار دارد و موتور جستجوی یاهو نیز با اختلاف بسیار کمی بعد از داک‌داک‌گو، کمترین میزان دقت را در این بررسی از خود نشان داد. علاوه بر آن، در این نمودار مشخص شده است که موتور جستجوی بینگ در جستجوی عبارات تیپ ۲ که عبارات پیچیده محسوب می‌شوند بیشترین میزان دقت را از خود نشان داده است. بعداز آن به ترتیب موتور جستجوی داک‌داک‌گو، یاهو و کلاز قرار دارند. به‌طورکلی، میانگین‌های به دست آمده از جستجوی عبارتی در این موتورهای جستجو نشان داد موتور جستجوی بینگ، داک‌داک‌گو، کلاز و یاهو به ترتیب بیشترین رکوردهای مرتبط را بازیابی کردند.

جدول ۲. میزان میانگین کل دقت به دست آمده برای هر دودسته موتور جستجوی نمونه

ردیف	موتورهای جستجو	میزان میانگین کل دقت (درصد)
۱	موتورهای جستجوی معنایی	۵۶/۱۴۵
۲	موتورهای جستجوی کلیدواژه‌ای	۵۷/۷۵

میزان میانگین کل دقت به دست آمده برای هر دودسته موتور جستجوی موردبررسی در جدول ۲ آمده است. همان‌طور که در این جدول مشخص شده است موتورهای جستجوی کلیدواژه‌ای نسبت به موتورهای جستجوی معنایی از میزان دقت بیشتری برخوردار هستند.

۵. بحث و نتیجه‌گیری

در پژوهش حاضر با محاسبه ضریب دقت به مقایسه عملکرد بازیابی اطلاعات موتورهای جستجوی معنایی و کلیدواژه‌ای بر اساس جستجوی عبارتی پرداخته شد. یافته‌ها نشان داد در جستجوی عبارات ساده، موتور جستجوی بینگ بیشترین مدارک مرتبط را بازیابی کرده است و بهترین عملکرد را از خود نشان داد. موتور جستجوی بینگ قابلیت تصمیم‌گیری و ارائه پاسخ فوری با پشتیبانی چندزبانه دارد و نوع نتایج ارائه‌شده در آن متن آزاد و لینک‌ها هستند (نگی و کومار، ۲۰۱۴). این موتور جستجو از زبان برنامه‌نویسی ASP. NET استفاده می‌کند و از زبان طراحی مترو[1] پیروی می‌کند و قابلیت نمایش نتایج مرتبط با جستجوی کاربر در سمت چپ صفحه نتایج[2] در آن وجود دارد. ویژگی منحصربه‌فرد این موتور جستجو «انجمن[3]» می‌باشد که در آن امکان یادگیری و به اشتراک‌گذاری منابع با افراد دیگری که از بینگ استفاده می‌کنند برای کاربران این موتور جستجو میسر است (دوتا و بانسال[4]، ۲۰۱۶). بعد از بینگ، موتور جستجوی داک داک گو بهترین عملکرد را داشت. داک داک گو با پاسخ‌های فوری (ناندی و دیگران[5]، ۲۰۱۶) و ارائه نتایج به‌صورت خلاصه‌های طبقه‌بندی‌شده به بازیابی اطلاعات می‌پردازد. این نتایج بر اساس منابع زیادی ازجمله یاهو، ویکی‌پدیا، والفرام آلفا و وب خزنده خودش جمع‌آوری می‌شود. علاوه بر آن، پشتیبانی چندزبانه دارد و رویکرد اصلی در آن جستجوی خوشه‌ای و پردازش زبان طبیعی[6] است (نگی و کومار، ۲۰۱۴). بعد از داک داک گو نیز موتور جستجوی کلاز قرار دارد. در کلاز نتایج ارائه‌شده شامل نمودار خوشه‌ای، تصاویر استخراج‌شده از داخل نتایج بازیابی شده، برچسب‌های لینک و نمودار معنایی بازیابی شده هستند. همچنین در این موتور جستجوی معنایی امکان نمایش نتایج مرتبط با جستجوی کاربر در سمت چپ صفحه نتایج وجود دارد. موتور جستجوی یاهو نیز در این مقایسه با بازیابی کمترین میزان رکوردهای مرتبط ضعیف‌ترین عملکرد را از خود نشان داد. این موتور جستجو که از نوع موتورهای جستجوی راهنما است (داورپناه، ۱۳۹۰، ۹۶) ساختاری سلسله مراتبی دارد و محتوای اصلی آن بر اخبار مالی متمرکز است (دوتا و بانسال، ۲۰۱۶). نتایج جستجو در آن به همراه توصیفی کوتاه از وب‌سایت‌های بازیابی شده ارائه می‌شود (داورپناه، ۱۳۹۰، ۹۶) و در آن امکان جستجوی مبتنی بر انتخاب وجود دارد (رات و دیگران[7]، ۲۰۱۳). همچنین، یافته‌ها نشان داد موتور جستجوی بینگ در جستجوی عبارات تیپ ۲ که عبارات پیچیده محسوب می‌شوند بهترین عملکرد را داشته است. بعدازآن به ترتیب موتور جستجوی داک‌داک‌گو، یاهو و کلاز قرار دارند. در این بررسی مشخص شد در جستجوی دو تیپ عبارات در نظر گرفته شده، موتور جستجوی بینگ بهتر از یاهو و موتور جستجوی داک داک گو بهتر از کلاز عمل کرده است. به‌طورکلی، میانگین‌های به دست آمده

1 . Metro
2. Search Engine Results Page (SERP)
3. Community
4 . Dutta & Bansal
5 . Nandy, Gohel, Kalyani, & Khakhariya
6. Natural Language Processing (NLP)
7 .Rathee, Atri, Rathee, & Nath

از جستجوی عبارتی در این موتورهای جستجو نشان داد موتور جستجوی کلیدواژه‌ای بینگ قوی‌تر از دو موتور جستجوی معنایی داک‌داک‌گو، کلاز و موتور جستجوی یاهو عمل کرده است. این نشان می‌دهد موتورهای جستجوی معنایی داک داک‌گو و کلاز بااینکه از استانداردهای وب معنایی استفاده می‌کنند اما نسبت به موتور جستجوی بینگ جستجوی عبارتی ضعیفی دارند. همچنین بر اساس این یافته‌ها می‌توان گفت موتور جستجوی بینگ به سمت معنایی شدن پیش می‌رود. با توجه به یافته‌های به دست آمده پیشنهاد می‌شود کاربران برای انجام جستجوی عبارتی به موتور جستجوی بینگ مراجعه نمایند.

اضافه بر این، موتورهای جستجوی معنایی بااینکه مدعی هستند که الف: توانایی درک پرس‌وجوهای کاربران را دارند؛ ب. متوجه می‌شوند که معنای نهفته در پرس‌وجو و مقصود کاربر چیست؛ پ. توانایی پاسخگویی به پرس‌وجوهای طولانی و پیچیده را دارند؛ ت. توانایی تطابق مفاهیم، تعمیم، متن‌کاوی در صفحات وب را دارند؛ ج. قادر به مقایسه لینک‌ها توسط آردی‌اف[1] و ارائه الگوهای گراف هستند؛ د. از قواعد استنتاج، منطق و سیستم زبان طبیعی بهره می‌برند (راماچانداران و سوجاتا، ۲۰۱۱). در این بررسی مشخص شد در جستجوی عبارتی نسبت به موتورهای جستجوی کلیدواژه‌ای برتری ندارند. در مقایسه‌ای که انجام گرفت، این موتورهای جستجو عملکرد قابل قبولی به‌عنوان موتورهای جستجوی وب معنایی از خود نشان ندادند و به نظر می‌رسد باید مسیر طولانی را طی نمایند تا به موتورهای جستجوی معنایی واقعی تبدیل شوند. برای نیل به این مهم لازم است از امکانات، ابزارها، ماژول‌ها، و فناوری‌های نوظهور عصر جدید از قبیل یادگیری ماشین، یادگیری عمیق، تلفیق این ماژول‌ها با تکنیک‌های متن‌کاوی، داده‌کاوی، و... بهره گیرند.

یافته‌های حاصل از پژوهش حاضر با نتایج پژوهش سینگ و شاران (۲۰۱۳)، احمدخان، سنگروا و رحمان (۲۰۱۴) و نگی و کومار (۲۰۱۴) همخوان است. در این پژوهش‌ها ده پرسش مورد جستجو قرار گرفت و بیست نتیجه اول بر اساس منطق کلاسیک موردبررسی قرار گرفت و مشخص شد که موتورهای جستجوی بینگ، داک داک گو و یاهو به ترتیب قوی‌ترین عملکرد را داشتند. اما نتایج پژوهش مالا و لوبیال (۲۰۱۶) با نتایج پژوهش حاضر ناهمخوان است. در این پژوهش نیز ده پرسش مورد جستجو قرار گرفت و بیست نتیجه اول بر اساس منطق کلاسیک موردبررسی قرار گرفت و مشخص شد موتور جستجوی داک داک گو، یاهو و بینگ به ترتیب بهترین عملکرد را از خود نشان دادند. به‌طورکلی، تعداد پرس‌وجوهای انتخاب‌شده برای ارزیابی عملکرد، تعداد رکوردهای بررسی شده حاصل از پرس‌وجوهای انجام‌شده، موضوعات جستجو شده در موتورهای جستجو، امکانات و ابزارهای جدید بکار گرفته شده توسط موتورهای جستجو و از همه مهم‌تر گذر زمان و خاصیت پویا و غیر ایستای وب باعث می‌شود موتورهای جستجو در زمان‌های متفاوت

1 . Rosource Description Framework (RDF)

عملکرد قوی یا ضعیفی داشته باشند. پژوهش حاضر نیز با تأثیر از این عوامل با چندی از پژوهش‌های پیشین همخوان و با برخی از آن‌ها ناهمخوان است.

٦. منابع و مآخذ

داورپناه، محمدرضا (۱۳۹۰). جستجوی اطلاعات علمی و پژوهشی در منابع چاپی و الکترونیکی. تهران: چاپار.

دری، راحله (۱۳۹۳). مقایسه و ارزیابی موتورهای جستجوی معنایی. *پردازش و مدیریت اطلاعات*، ۳۰ (۲)، ۴۶۷-۴۸۷.

شیری، علی اصغر (۱۳۷۳). ارزیابی بازیافت و دقت نظام های بازیابی اطلاعات. *تحقیقات اطلاع رسانی و کتابخانه های عمومی (پیام کتابخانه سابق)*، ۴(۱و۲)، ۳۸- ۴۵.

کوشا، کیوان (۱۳۸۱). ابزارهای کاوش اینترنت: اصول، مهارتها و امکانات جستجو در وب. تهران: کتابدار.

لنکستر، اف دبلیو (۱۳۸۲). نمایه سازی و چکیده نویسی: مبانی نظری و عملی. (مترجم: عباس گیلوری). تهران: چاپار.

نادی راوندی، سمیه ؛و نحریری، نجلا (۱۳۹۵). نظام های بازیابی اطلاعات. تهران: کتابدار.

نوروزی، مرتضی ؛و طاهریان، محسن (۱۳۹۰). وب معنایی. تهران: فرس.

Khan, J.A., Sangroha, D., Ahmad, M., & Rahman, M.T. (2014). A performance evaluation of semantic based search engines and keyword based search engines. 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 168-173. Retrieved 23 October, 2017 from: http://ieeexplore.ieee.org/document/7005997/

Aldabbagh, M. S. M. (2012). An Approach for Evaluation of Semantic Performance of Search Engines: Google, Yahoo, Msn and Hakia. *J. Edu. & Sci*, 25: 124-130.

Aliyu, F. M., & Yahaya, Y. I. (2021). An Investigation of the Accuracy of Knowledge Graph-base Search Engines: Google knowledge Graph, Bing Satori and Wolfram Alpha. International Journal of Scientific & Engineering Research, 12(1), 11-15.

Andago, M.0., Phoebe, T.P., & Thanoun, B.A. (2011). Evaluation of a Semantic Search Engine against a Keyword Search Engine Using First 20 Precision 1.

Azizan, A., Abu Bakar, Z., Abd Rahman, N., Masrom, S. & Khairuddin, N. (2018). A Comparative Evaluation of Search Engines on Finding Specific Domain Information on the Web. International Journal of Engineering & Technology, 7 (4), 1-4.

Choudhury, N. (2014). World Wide Web and Its Journey from Web 1.0 toWeb 4.0. Nupur Choudhury/ (IJCSIT). International Journal of Computer Science and Information Technologies, 5 (6),8096-80100.

Dutta, M., & Bansal, K. L. (2016). A Review Paper on Various Search Engines (Google, Yahoo, Altavista, Ask and Bing). International Journal on Recent and Innovation Trends in Computing and Communication, 4 (8),190- 195.

Fatima, A., Luca, C., & Wilson, G. (2014). New Framework for Semantic Search Engine. International Conference on Computer Modelling and Simulation. UKSim-AMSS,445-450. DOI:10.1109/UKSim.2014.114

Gayar, M.M.E., Mekky N., & Atwan, A. (2015). Efficient Proposed Framework for Semantic Search Engine using New Semantic Ranking Algorithm. (IJACSA) International Journal of Advanced Computer Science and Applications, 6(8),136 - 143.

Goutam, R. K., & Dwivedi, S. K. (2012). Performance Evaluation of search engines via user efforts Measures. *International Journal of Computer Science Issues*, 9(2),432- 437.

Guha, R., McCool, R. ,& Miller, E. (2003). *Semantic Search*. Retrieved 23 October, 2017 from: https://dl.acm.org/citation.cfm?id=775250

Signorini, A., & Imielinski, T. (2009). If You Ask Nicely, I will Answer: Semantic Search and Today's Search Engines. 2009 IEEE International Conference on Semantic Computing, 184-191.

Kumar, D., & Mishra, R. (2015). Deep Web Performance Enhance on Search Engine. *2015 International Conference on Soft Computing Techniques and Implementations- (ICSCTI), Hybrid Intelligence*: 137- 140.

Mala, V., & Lobiyal, D. K. (2016). Semantic and keyword based web techniques in information retrieval. 2016 International Conference on Computing, Communication and Automation (ICCCA). https://doi.org/10.1109/ccaa.2016.7813724

Malve, A., & Chawan, P. M. (2015). A Comparative Study of Keyword and Semantic based Search Engine. *International Journal of Innovative Research in Science,Engineering and Technology*, 11(4),11156-11161.

Nandy, D., Gohel, H., Kalyani, F. ,& Khakhariya, M. (2016). DuckDuckGo vs. Google search engine comparative analysis. International Journal of Science Research and Technology, 2(1), 37-41.

Negi, Y. S., & Kumar, S. (2014). A Comparative Analysis of Keyword-and Semantic-Based Search Engines. In Intelligent Computing, Networking, and Informatics(pp.727-736).

Ramachandran, A., & Sujatha, R. (2011). Semantic search engine: A survey. Int. J. Comp. Tech. Appl, 2 (6),1806-1811.

Rashid, J., & Wasif Nisar, M. (2016). A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines. I.J. Information Technology and Computer Science, 10(10),82- 89.

Rathee, D., Atri, J., Rathee, J., & Nath, R. (2013). A Framework to Evaluate Search Engines. IJCSC, 4 (2), 184-190.

Sayed, A., & Al Muqrishi, A. (2017). IBRI-CASONTO: Ontology-based semantic search engine, Egyptian Informatics Journal,18 (3), 181- 192.

Sheela, A.C., & Jayakumar, C.T. (2019). Comparative Study of Syntactic Search Engine and Semantic Search Engine: A Survey. 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), 1, 1-4.

Singh, J., & Sharan, A. (2013). A Comparative Study between Keyword and Semantic Based Search Engines. *International Conference on Cloud, Big Data and Trust 2013*: 130- 134.

Solomou, G. ,& Koutsomitropoulos, D. (2015). Towards an evaluation of semantic searching in digital repositories: a DSpace case-study. *Electronic library and information systems*, 49 (1), 63-90.

Sudeepthi, G., Anuradha, G., Surendra, M.,& Babu, P. (2012). A Survey on Semantic Web Search Engine. *IJCSI International Journal of Computer Science Issues*, 9(2), 241-245.

Tümer, D., Shah, M.A., & Bitirim, Y. (2009). An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia. 2009 Fourth International Conference on Internet Monitoring and Protection, 51-55.

Yadav, A., & Maheshwari, S. (2017). Building of Intellectual Web Search Engines through Semantic Web. *International Journal of Latest Trends in Engineering and Technology*, 7 (3), 384-391.