



Investigating the Degree of Web Search Engine Efficiency in Retrieval of Information Related to the Information & Knowledge Domain Based on seven indicators

Elahe Bahmaee¹, Mohammad Reza Farhadpoor^{2*}

1. Department of Information Management, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
2. Department of Information Management, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

Article Info.

Received: 2018/09/30

Accepted: 2019/11/06

Abstract

Background and Objectives: The advent of World Wide Web (WWW) in 1990s that was followed by emergence of a large number of web pages made using of automatic information retrieval systems necessary. The first web search engine with a capability of full text search was developed by Brian Pinkerton in the Washington University. The Web Crawler was able to index the plain texts and allowed the users to search the expressions on the internet. Later, the Lycos, Infoseek in 1994, Excite and Yahoo in 1995 Inktomi in 1996, Google in Sep. 1997 (Gross, 2015), M.S.N and Overture (Sahu, Mahapatra and Balabantaray, 2016) were emerged to overcome the complexity resulted from a surge in the information within the web. According to Wu and Lee (2004), the services delivering by web search play a significant role for those users who seek to elicit information sources to meet their needs which had not been available for them before. Today, the search engines are recognized as an access pass to a huge size of information on the internet, providing the services and tools to meet a variety of users' information requirements. For this reason, the evaluation of efficiency and performance of search engines is very important because it is necessary for developers and users. (Azimzadeh, Badie and Esnaashari, 2016). As far as Mc Carthy (2006: quoted in Ewans, 2007) claims, a vast majority of the people visiting the web sites access to the webpages or contents of interest through search engines rather than link directly from other pages.

The search engine is a software by which the users search the needed various information on the internet as well as retrieval of related outcomes (Mivule, 2017). Craft, Metzler and Strohmman (2015) define a search engine as a practical use of techniques retrieving information within the large scale text sets that with different forms reflect those capabilities for which they are designed purposely. On the other word, the search engines are the programs that are employing to find the documents matched to the specific keywords on the WWW setting and retrieve a list of documents containing the searched keywords (Khorsheed, Madbouly and Guirguis, 2015). According to Craft, et. al (2015) the capabilities of

recovering information are dramatically involving in the structured multimedia documents, meaningful textual contents and other media, relevance, evaluation, information needs, the effective ranking algorithms and interaction with the users, what are still concerned the researchers investigating in retrieval of information. From Ali, Jhandir, Lee, On and Choi (2017) viewpoints, while the data performance for internet acts as a fuel to back running it, its extensiveness has caused much problems for the users.

While the degree of the users' confidence to the search engines and relying on it to display authentic outcomes is questionable, providing suitable, relevant, and high quality information for the users using webpages contents and links between pages is a big challenge for service providers (the search engines). It is while Xu, Wang and Goh (1998) believe that the numerous search engines have been developed to give technically better performance. It indicates that there has been lack of expected qualitative features from users' viewpoints.

Given that in the various studies a few and mostly specific measures such as precision and recall have been considered, on the basis of this assumption that a relevant document collection is ever the same without involvement of user, in the present study the relative precision, relative recall, F-number, coverage ratio, freshness ratio, the expected search ratio and failure are investigated as a set of measures evaluating the retrieval efficiency of the search engines in the information and knowledge domain. Therefore, the primary purpose of this study is to determine the retrieval efficiency of the five search engines given the indicators of interest. Meanwhile, the secondary purpose of this article is to identify the retrieval efficiency of the search engines based on such indicators as the relative precision, relative recall, F-number, coverage ratio, freshness ratio, the expected search ratio and failure. The main question is that in this domain how the web search engines operate efficiently to retrieve the information.

Methodology: The present study in term of target is applied one and is descriptive in term of survey method. By taking into account the search function and search term as identical condition (the query AND pdf), the authors have evaluated the efficiency of the search engines based on what is observed in the retrieval results. To measure the variables, several formulas related to relative precision, relative recall, F-number, coverage ratio, freshness ratio, the expected search ratio and failure have been used. Alexa-based ranking, this study also investigates such search engines as Google, Bing, Yahoo, Ask and AOL as the most commonly used search engines. In this research, the data was collected through library studies, in order to write a research literature, checklists and through direct observation of the searched outcomes. For this purpose, based on article, Top Trends in Academic Libraries 2106, published in the SAGE database, Collage Research Libraries journal, the new domains and future subfields in the information science were determined in 15 terms/ keywords. Then, the 15 terms/keywords were analyzed and to determine the search keywords they were assigned to 20 experts of the field to give a weighted score in order to be placed in the list of this study by a value given to them. Then, 5 keywords with higher weights were selected for searching. The 5 keywords were individually assigned to the 5 search engines each. In the next stage, out of total retrieved outcomes of each search engine for each keyword, 50 documents were placed on the top of the list to evaluate the retrieval efficiency measures in the researcher-made collection. In order to distinguish relevance of documents, according to Zhang, Xu, Wang and Lee (2006)' recommendations, the keyword iteration in the documents, abstracts and their keywords was considered as a measure. To assess the reliability of retrieved results, the retest method was used. For this purpose, over two phases, at 15-day interval (winter 2017) the search and

retrieval were conducted again. The correlation results of the two experiments were tested and confirmed at the $R=0.89$. To analyze the data, the Excel 2013 was employed.

Findings: The results showed that the search engines of Google, Ask and Yahoo are of better performance than the other search engines in term of relative precision, relative recall, F-number harmonic mean and failure criterion; the expected search coefficient indicators and freshness ratio; coverage ratio respectively. However, in spite of the results obtained for the study search engines based on the different indicators, in general they are not in ideal situation where in most cases they are lower than the average. Given that in the previously studied search engines, the scientific domain and indicators of efficiency were different than the present one and they have focused mostly on the relative precision and relative recall, it is not possible to compare all the findings. However, the findings from the present study based on indicator of relative precision are consistent with that of studies of Shafi and Rader (2005), Ali and Gole (2016) and on the recall with that of Janson and Molina (2006), Kumar and Prakash (2009), Wang et. Al (2012), Ali and Gole (2016) and on the indicator of coverage ratio with that of Mohammad Ismael and Mansoor Kiakie (2011), Esfandyari Moghaddam (2012) and Janson and Molina (2006).

Discussion: It can be concluded that in spite of the search engines have been gained a score and ranked in this study but they are still far from ideals. It follows that the challenges related to evaluation of information retrieval efficiency, despite of using different search engines, implementation of various strategies and different ranking algorithms and methods of document indexing, has to be yet removed and they need more studies. It should be noted that the comparison of the results from this study with some prior findings indicates that neither of search engines can alone meet the required efficiency indicators. Thus, given the different indicators the users have to assign their queries to the search engines. On the one hand, designing the specific search engines with regard to diversity, extent and lexical relationship in the different domains of sciences is very necessary. On the other hand, it is the time to use various patterns such as visual searching, using multilingual thesaurus, retrieving based on weighed indexing in the interface of both specific and general search engines.

Keywords: *Internet Search Engines, Retrieval performance, Information Science and Knowledge, Information Retrieval, Web Search Engines.*

***Corresponding author:**

Email: M.farhadpoor@gmail.com

How to Cite:

Bahmaee, E., Farhadpoor, M. (2020). Investigating the Degree of Web Search Engine Efficiency in Retrieval of Information Related to the Information & Knowledge Domain Based on seven indicators. *Journal of Studies in Library and Information Science*, 12(2): 31-48.



بررسی میزان کارآیی موتورهای کاوش وب در بازیابی اطلاعات مربوط به حوزه اطلاعات و دانش به لحاظ شاخصه‌های هفت گانه

الهه بهمنی^۱، محمدرضا فرهادپور^{۲*}

۱. گروه مدیریت اطلاعات، واحد اهواز، دانشگاه آزاد اسلامی، اهواز ایران

۲. گروه مدیریت اطلاعات، واحد اهواز، دانشگاه آزاد اسلامی، اهواز ایران

اطلاعات مقاله

تاریخ پذیرش: ۹۸/۰۸/۱۵

تاریخ دریافت: ۹۷/۰۷/۰۸

چکیده

هدف: هدف مطالعه بررسی میزان کارآیی موتورهای کاوش وب در بازیابی اطلاعات در حوزه اطلاعات و دانش است. **روش‌شناسی:** روش تحقیق پیمایشی توصیفی است. تعداد ۵ کلیدواژه از مقاله ۲۰۱۶ top trends in academic libraries در مجله College and Research Libraries با رتبه‌بندی ۲۰ متخصص رشته انتخاب شد. سپس ۵۰ مدرک بازیابی شده ابتدای لیست پنج موتور کاوش گوگل، یاهو، اسک، بینگ و ای.او.ال. در گردآورد قرار گرفت و با رعایت فرمول جستجو یکسان، کارآیی موتورهای کاوش بررسی شد.

یافته‌ها: بر اساس یافته‌ها کارایی گوگل از نظر شاخصه‌های مانعیت، ضریب F و نسبت ناکامی به ترتیب با مقادیر ۰/۸۲۴، ۰/۳۶۰ و ۸/۸؛ و اسک با مقادیر ۰/۳۷۸ و ۰/۶۰ از نظر شاخصه‌های نسبت تازگی و ضریب کاوش مورد انتظار در سطح بالاتری هستند. ضمناً گوگل و اسک به ترتیب با مقدار جامعیت نسبی ۰/۲۴ و ۰/۲۲ و در رابطه با شاخصه نسبت پوشش، یاهو با ۰/۷۱۴ وضعیت بهتری دارد.

نتیجه‌گیری: برای دقت بازیابی موتور کاوش گوگل، برای اطلاعات جدید موتور کاوش اسک و برای جامعیت بازیابی موتور کاوش گوگل همراه با یاهو دارای کارایی قابل قبول هستند.

کلیدواژه‌ها: موتورهای کاوش اینترنت، کارآیی بازیابی، علم اطلاعات و دانش، بازیابی اطلاعات، موتورهای کاوش وب

*نویسنده مسئول: M.farhadpoor@gmail.com

استناد به این مقاله:

بهمنی، الهه، فرهادپور، محمدرضا (۱۳۹۹). بررسی میزان کارآیی موتورهای کاوش وب در بازیابی اطلاعات مربوط به حوزه اطلاعات و دانش به لحاظ شاخصه‌های هفت گانه. مطالعات کتابداری و علم اطلاعات، ۱۲ (۲): ۳۱-۴۸

مقدمه و بیان مسئله

ظهور وب جهان گستر در دهه ۱۹۹۰ و به تبع آن افزایش تعداد صفحات وب، وجود نظام‌های بازیابی اطلاعات خودکار را ضرورت بخشید. نخستین موتور کاوش وب با قابلیت پیمایش تمام متن در ۱۹۹۴ توسط برایان پینکرتون^۱ در دانشگاه واشنگتن عرضه شد. وب کراولر قادر بود متون ساده را نمایه کند و به کاربران اجازه جستجوی اطلاعات را در اینترنت می‌داد. در ادامه لایکوز^۲ اینفوسیک^۳ در ۱۹۹۴، اکسایت^۴ آلتاویستا^۵ و یاهو در ۱۹۹۵، اینکتومی^۶ در ۱۹۹۶، گوگل در سپتامبر ۱۹۹۷ (گروس^۷؛ ۲۰۱۵)، ام.اس.ان و آورتور^۸ در ۱۹۹۸ (ساهو^۹، مهاپاترا^{۱۰} و بالابانتارای^{۱۱}؛ ۲۰۱۶) پدید آمدند که همگی با هدف فائق آمدن بر پیچیدگی ناشی از رشد سیل آسای اطلاعات وبی بود. وو^{۱۲} و لی^{۱۳} (۲۰۰۴) اینترنت و وب جهان گستر را یکی از مصنوعات مهم و قابل تامل نسل بشر معرفی می‌کنند که از نخستین ظهورشان در ۱۹۹۳ تاکنون در هر جایی رواج یافته است و از بین کاربردها و خدمات متنوع آن بازیابی اطلاعات و ایمیل پرکاربردترین قابلیت‌های آن بوده است. همزمان با بلوغ وب جهان گستر، موتورهای کاوش به شکل قابل توجهی جایگاه قدرتمندی بدست آوردند که به خاطر قابلیت آنها در کانالیزه کردن توجه میلیون‌ها کاربر و ایجاد منافع برای وب سایت‌ها از طریق برنامه‌های تبلیغاتی بافتاری بود (اونز^{۱۴}؛ ۲۰۰۷). در همین ارتباط لوین^{۱۵} (۲۰۱۰) نقش آفرینی موتورهای کاوش در تلاش آنها برای رسیدن به پاسخ (یعنی اطلاعات) یک سوال ویژه (یعنی پرس‌وجو) را نمونه ساده‌ای از کسب این جایگاه در نگاه کاربران می‌داند.

به اعتقاد وو و لی (۲۰۰۴) خدمات جستجوی وب برای کاربرانی که به دنبال تشخیص منابع اطلاعاتی مرتبط با نیازی خاص هستند که پیش از آن ناشناخته بودند، نقش مهمی ایفا می‌کند. امروزه موتورهای کاوش به عنوان گذرگاه دسترسی به حجم عظیمی از اطلاعات در اینترنت باز شناخته شده اند که خدمات و ابزارهایی را برای حل نیازهای اطلاعاتی متنوع کاربران فراهم می‌سازند. برای همین ارزیابی کارآیی و عملکرد موتورهای کاوش به جهت اهمیتی که برای توسعه‌دهندگان و کاربران دارد، ضروری است (عظیم‌زاده^{۱۶}، بدیع^{۱۷} و اثنی‌عشری^{۱۸}؛ ۲۰۱۶). مک کارتی^{۱۹} (۲۰۰۶) نقل در: اونز (۲۰۰۷) مدعی است، بیش از نیمی از بازدیدکنندگان وب سایت‌ها به جای پیوند مستقیم از سایر صفحات، از طریق موتورهای کاوش به محتوای/صفحه

1 Brian Pinkerton

2 WebCrawler

3 Lycos

4 Infoseek

5 Excite

6 AltaVista

7 Nkntomi

8 Gross

9 Overture

10 Sahu

11 Mahapatra

12 Balabantaray

13 Wu

14 Li

15 Evans

16 Eevene

17 Azimzadeh

18 Badie

19 Snaashari

20 McCarthy

مورد نظر دست پیدا می‌کنند. موتور کاوش ابزاری نرم‌افزاری برای جستجوی اطلاعات با موضوعات مختلف در اینترنت و بازیابی نتایج مرتبط به کاربران است (از نظر میول، ۲۰۱۷). کرافت، متزلر^۳ و استرومن^۴ (۲۰۱۵) موتور کاوش را کاربرد عملی فنون بازیابی اطلاعات در مجموعه‌های متنی مقیاس بزرگ می‌دانند که با اشکال متنوع، بازتاب قابلیت‌هایی هستند که به آن منظور طراحی شده‌اند. به بیان دیگر، موتورهای کاوش برنامه‌هایی هستند که برای پیدا کردن اسناد و مدارک مرتبط با کلیدواژه‌های خاص در محیط وب جهان گستر بکار می‌روند و فهرستی از مدارک حاوی کلیدواژه‌های جستجو شده را بازیابی می‌کنند (خورشید، مذبولی^۵ و گورگیس^۶، ۲۰۱۵). به اعتقاد کرافت و همکاران (۲۰۱۵) قابلیت‌های بازیابی اطلاعات به طرز چشم‌گیری درگیر مدارک چندرسانه‌ای ساختارمند، محتوای متنی معنی‌دار، و سایر رسانه‌هاست و ربط، ارزیابی و نیازهای اطلاعاتی، الگوریتم‌های رتبه‌بندی موثر و تعامل با کاربران مسائلی کلیدی هستند که در عصر موتورهای کاوش هنوز دغدغه اصلی محققان بازیابی اطلاعات باقی مانده است. از دیدگاه علی، ژاندریر^۷، لی^۸، آن^۹ و چوی^{۱۰} (۲۰۱۷) در حالی که عملکرد داده‌ها برای اینترنت به مانند سوختی است که به راه‌اندازی آن کمک می‌کند، اما فراوانی آن در وب جهان گستر سرمنشاء مشکلات زیادی برای کاربران است. این در حالی است که حدود نیمی از جمعیت جهان از اینترنت استفاده می‌کنند و بخش عمده‌ای از آنها از موتورهای کاوش استفاده می‌کنند.

کولکارنی^{۱۱} (۲۰۱۳) دسترسی روزافزون به مجموعه‌های بزرگ اطلاعاتی و پیشرفت الگوریتم‌های بازیابی در سیستم‌های مدرن جستجو را از جمله دلایل وابستگی بیش از حد کاربران به جستجو معرفی می‌کند و وب را نمونه عینی این استدلال می‌داند که استخراج اطلاعات از این مجموعه بزرگ را برای افراد و سازمان‌ها ارشمند ساخته است. براساس پیش‌بینی‌های مطالعات اخیر تا سال ۲۰۲۰ بالغ بر ۱۰ میلیارد ابزار و حسگر وجود خواهد داشت که بالغ بر ۴۰ زینتابایت معادل ۴۰ تریلیون گیگابایت داده تولید خواهند کرد. این ابزارهای همبسته^{۱۲} در محیط‌های بلادرنگ^{۱۳} مستقر و در ارتباط با گردآوری، ارتباط و توزیع داده‌ها نقش‌آفرینی خواهند کرد (علی و همکاران، ۲۰۱۷). کاتومبا^{۱۴} و کوتزه^{۱۵} (۲۰۱۷) نیز، با استناد به نتایج پروژه اینترنت پیو^{۱۶} آمریکا، مدعی است که بالغ بر ۹۲ درصد از فعالیت‌های اینترنتی درگیر استفاده از موتورهای کاوش است و حدود ۳ میلیارد جستجو که در سال ۲۰۰۹ به گوگل سپرده می‌شد تا سال ۲۰۱۶ از مرز ۵/۵ میلیارد جستجو در روز گذشته است. از نظر سیویه^{۱۷} و دهبینو^{۱۸} (۲۰۱۶) امروزه وب برای بسیاری از مردم جهان منبع اصلی اطلاعات است و نقش غالبی را به عنوان

^۱Mivule

^۲Croft

^۳Metzler

^۴Strohman

^۵Khorsheed

^۶Madbouly

^۷Guirguis

^۸Jhandir

^۹Lee

^{۱۰}On

^{۱۱}Choi

^{۱۲}Kulkarni

^{۱۳}Interconnected

^{۱۴}Katumba

^{۱۵}Coetsee

^{۱۶}Pew Internet Project (PIP)

^{۱۷}Sibuyi

^{۱۸}Dehinbo

یک سازوکار در انتشار و بازیابی اطلاعات در اینترنت ایفا می‌کند. در حالی که میزان اعتماد کاربران به موتورهای کاوش و اتکا به آن جهت نمایش نتایج معتبر جای تامل دارد، ارائه اطلاعات مناسب، مرتبط و با کیفیت به کاربران با استفاده از محتوای صفحه وب و پیوند میان صفحات چالش بزرگی برای ارائه دهندگان خدمات (موتورهای کاوش) به حساب می‌آید. این در حالی است که به اعتقاد ژئی، وانگ^۳ و گو^۳ (۱۹۹۸) موتورهای کاوش متعدد با هدف عملکرد فنی بهتر توسعه یافته‌اند که نشان می‌دهد از دیدگاه کاربران فاقد ویژگی‌های کیفی مورد انتظار بوده‌اند.

توسعه موتورهای کاوش با عنایت به محدودیت‌های فناوری‌های وب با هدف غلبه بر چالش‌هایی از قبیل ساختار محتوای وب برای بازیابی اطلاعات، ابهام ناشی از فقدان پیوند و همبستگی درونی بین اطلاعات، نقص در انتقال خودکار اطلاعات، کند بودن در خصوص مجموعه‌های بزرگ کاربران و محتوا، و نبود قالب جهانی برای نظام‌ها از منظر درک اطلاعات بوده است (کرافت و همکاران، ۲۰۱۵). ساهو و همکاران (۲۰۱۶) نیز با تاکید بر تعریف سالتون از بازیابی اطلاعات که آن را حوزه‌ای علمی در ارتباط با ساختار، سازماندهی، تحلیل، ذخیره، جستجو و بازیابی اطلاعات معرفی می‌کند (۱۹۶۸) نقل در: کرافت و همکاران، ۲۰۱۵؛ ابهام^۴ محتوای ذهنی^۵ فرار بودن^۶ نرخ بالای تغییرات فناورانه، اعتبار نتایج، تاثیرات انگیزشی را از چالش‌های حوزه مطالعات موتورهای کاوش معرفی می‌کند. به نحوی که در خیلی از مواقع پرس‌وجوی کاربر ما را با حجم زیادی از مدارک بی‌ربط مواجه می‌سازد که ناتوانی موتورهای کاوش در تضمین بازیابی داده‌های مرتبط را نشان می‌دهد. به اعتقاد جانسن^۷ و مولینا^۸ (۲۰۰۶) اینترنت هنوز هم این قابلیت را دارد که منشاء شگفتی‌ها باشد و ویژگی‌های کاربران در کنار ماهیت پویای موتورهای جستجو از جمله ضرورت‌های ارزیابی موتورهای کاوش است. نتایج مطالعات گذشته (مانند بایفت^۹ کاستیلو^{۱۰}، چیریتا^{۱۱} و وبر^{۱۲}؛ موران^{۱۳} و هانت^{۱۴}؛ فورتناتو^{۱۵}، بوگونا^{۱۶}، فلمینی^{۱۷} و منزر^{۱۸}؛ ۲۰۰۶) نقل در اونز (۲۰۰۷) دو دسته عوامل مهم شامل عوامل مربوط به پرس‌وجو (که به محتوای صفحه وب مانند وجود و بسامد کلیدواژه‌ها متکی هستند) و عوامل مستقل از پرس‌وجو (به اطلاعات صفحات وب خارجی مربوط است که به صفحه مورد نظر پیوند داده‌اند) را شناسایی کرده‌اند که به طور بالقوه قابلیت موتورهای کاوش در رتبه بندی یک وب‌سایت را تحت تاثیر قرار می‌دهند. از دیدگاه شاکری (۱۳۸۷) نیز هزینه‌های جستجو افزایش یافته است و زمان قابل توجهی برای جستجوی اطلاعات در پایگاه‌ها توسط کاربران صرف می‌شود که ارزیابی کارآیی موتورهای جستجوی مختلف در بازیابی اطلاعات را ضرورت می‌بخشد.

Xie

Wang

Goh

Ambiguity

Subjective Content

Volatility

Jansen

Molina

Bifet

Castillo

Chirita

Weber

Moran

Hunt

Fortunato

Boguna

Flammini

Menczer

با توجه به این که در مطالعات مختلف عمدتاً معیارهای خاص و محدودی مانند جامعیت و مانعیت مورد توجه بوده‌اند که توجه آنها بر این فرض متمرکز است که مجموعه اسناد مرتبط برای یک پرسش، در همه حال یکسان و مستقل از کاربر است؛ در این مطالعه، جامعیت نسبی، مانعیت نسبی، ضریب F، نسبت پوشش، نسبت تازگی، ضریب کاوش مورد انتظار و ناکامی به عنوان مجموعه‌ای از سنج‌ها و معیارهای ارزیابی کارایی بازیابی موتورهای جستجو در موضوعات حوزه اطلاعات و دانش مورد بررسی قرار گرفت. لذا با هدف تعیین کارایی بازیابی موتورهای کاوش پنج‌گانه با توجه به شاخصه‌های مذکور، سوال اصلی پژوهش حاضر این است که کارایی موتورهای کاوش وب در بازیابی اطلاعات حوزه اطلاعات و دانش چگونه است؟

پیشینه‌های پژوهش

موضوع ارزیابی کارایی و عملکرد موتورهای کاوش به لحاظ ناتوانی آنها در ارائه پاسخ‌های مرتبط به پرس‌وجوی کاربران در قالب مدارک بازیابی شده به نحو مطلوب، مورد توجه مطالعات مختلفی بوده است که با توجه به تعدد متغیرهای دخیل در کارایی و عملکرد موتورهای کاوش و نتایج بازیابی اطلاعات از قبیل شیوه‌های نمایه‌سازی، الگوریتم‌های رتبه‌بندی مختلف، مسایل زبانی، ذهنی و وابسته به کاربر بودن ربط، تنوع نیازهای اطلاعاتی، ویژگی‌های نرم‌افزاری موتورهای کاوش و غیره، هر کدام از زاویه خاص به مساله پرداخته‌اند. به اعتقاد کوشا (۱۳۸۲) طراحی و توسعه یک ابزار کاوش ملی با امکانات و قابلیت‌های جستجوی پیشرفته به زبان فارسی در داخل کشور ضروری است. نتایج مطالعه شاکری (۱۳۸۷) با هدف ارزیابی جامعیت و مانعیت ابزارهای کاوش فارسی اینترنت در بازیابی اطلاعات در حوزه کتابداری و اطلاع‌رسانی نشان داد که بین مانعیت و نیز جامعیت ابزارهای کاوش، اختلاف معناداری وجود دارد. نتایج مطالعه صراطی شیرازی (۱۳۸۸) با عنوان مقایسه میزان دقت موتورهای کاوش عمومی و تخصصی پزشکی در بازیابی مدارک مربوط به بیماری‌های کودکان، نشان داد از نظر دقت و ریزش کاذب، موتورهای کاوش عمومی تفاوت معناداری با هم ندارند. اما این تفاوت در میان موتورهای کاوش تخصصی پزشکی معنی‌دار بوده و کم‌ترین دقت و بیشترین ریزش کاذب مربوط به موتور کاوش مدولوژی^۱ و کم‌ترین ریزش کاذب مربوط به موتور کاوش تریپ‌دیتابیس^۲ بود. هم‌چنین میزان دقت موتورهای کاوش عمومی و تخصصی پزشکی با یکدیگر اختلاف معناداری دارد و نکته جالب این که، دقت در موتورهای کاوش عمومی بیشتر از موتورهای کاوش تخصصی پزشکی است. براساس نتایج پژوهش محمداسماعیل و منصورکیایی (۱۳۹۰) با موضوع مقایسه موتورها و ابرموتورهای کاوش عمومی در بازیابی اطلاعات علم فیزیک و میزان هم‌پوشانی آنها، موتور کاوش یا هو بیشترین مدارک فیزیک (۴۰٪) را بازیابی کرد. ضمن این که نتایج موتور کاوش عمومی ای.اوال. با حدود (۳۸/۵۷ درصد) بیشترین هم‌پوشانی را با نتایج سایر موتورهای کاوش دارد. ابرموتور کاوش کری گاید نیز بیشترین مدارک فیزیک (۷۷/۱ درصد) را بازیابی کرد. هم‌چنین ابرموتور اینفو با (۴۳/۷ درصد) بیشترین میزان هم‌پوشانی را با سایر ابرموتورها داشت. در نهایت محققان پیشنهاد می‌کنند که استفاده از ابرموتور کاوش عمومی نسبت به موتور کاوش عمومی بهتر است. هم‌چنین در پژوهش دیگری اسفندیاری مقدم (۱۳۹۱) به بررسی هم‌پوشانی نتایج بازیابی شده کلیدواژه‌های تخصصی پزشکی در موتورهای کاوش عمومی وب پرداخت. بر اساس یافته‌های این مطالعه موتور کاوش یا هو دارای بیشترین نتیجه‌ای بود که تنها در خود آن بازیابی شده بود.

^۱Medology

^۲Trip Database

در حوزه مطالعات گذشته نیز، نتایج مطالعه شافی^۱ و رادر^۲ (۲۰۰۵) با موضوع جامعیت و مانعیت پنج موتور کاوش گوگل، آلتاویستا، سایروس، هات بوت، بایوبوب را در بازایی اطلاعات علمی حوزه بیوتکنولوژی با استفاده از عبارات جست و جوی مرتبط با بیوتکنولوژی و بر اساس انتخاب ده نتیجه اول بازایی شده، نشان داد عملکرد سایروس در بازایی مدارک علمی بهتر است و نیز بالاترین میزان جامعیت و مانعیت را کسب کرده است. گوگل بهترین جانشین برای بازایی مدارک علمی وبی است. هات بات ترکیب خوبی از جامعیت و مانعیت را ارائه داد اما هم پوشانی بیشتری با موتورهای کاوش دیگر داشت. بر اساس یافته‌ها مانعیت و جامعیت رابطه معکوسی با یکدیگر دارند. نتایج مطالعه جانسن و مولینا (۲۰۰۶) با هدف ارزیابی کارایی موتورهای کاوش در بازایی پیوندهای مرتبط با تجارت الکترونیکی، نشان داد که پیوندهای بازایی شده به واسطه موتورهای جستجوی تجارت الکترونیکی در مقایسه با سایر موتورهای کاوش وضعیت بهتری دارند ولی تفاوت معناداری با عملکرد خدمات دایرکتوری‌های وب نداشت. از نظر مانعیت موتور کاوش گوگل بهترین عملکرد را داشته است و پس از آن بینگ و یاهو و اسک به ترتیب در رتبه‌های بعدی قرار داشته‌اند. ضمن این که یاهو دایرکتوری و فروگل در مقایسه با اکسایت، گوگل و آورتور مدارک مرتبط بیشتری را بازایی کردند. نتایج مطالعه سمپات کومار^۳ و پراکاش^۴ (۲۰۰۹) با عنوان مطالعه مقایسه‌ای مانعیت و جامعیت نسبی گوگل و یاهو با بکارگیری پرس وجوهای تک‌واژه‌ای، چند واژه‌ای ساده و چندواژه‌ای پیچیده از حوزه کتابداری و اطلاع‌رسانی نشان داد که دقت گوگل برای پرس وجوهای چندواژه‌ای ساده با نسبت ۰/۹۷ بالاست و در ارتباط با پرس وجوهای چندواژه‌ای پیچیده دقت یاهو با نسبت ۰/۷۶ بالاست. به لحاظ جامعیت نسبی نیز کارایی گوگل برای پرس وجوهای تک‌واژه‌ای ساده با نسبت ۰/۹۲ وضعیت مطلوبی دارد در حالی که به لحاظ همین شاخصه وضعیت یاهو برای پرس وجوهای چندواژه‌ای با نسبت ۰/۶۱ بهتر است. ونگ^۵ و همکاران (۲۰۱۲)، نیز در مطالعه خود چهار موتور کاوش اصلی بینگ، یاهو، گوگل و اسک را برای دستیابی اطلاعات پزشکی را بر اساس نتایج ۲۰۰ پیوند اول هر موتور کاوش مورد بررسی قرار دادند. نتایج به دست آمده نشان داد که گوگل بیشترین اعتبار را دارد. ضمن این که نتایج به دست آمده از موتورهای کاوش هم‌چنین تا حد زیادی با هم مشترک بودند. در مطالعه دیگری، دوماچوفسکی^۶، گریسباوم^۷ و هیوینگ^۸ (۲۰۱۵) با اشاره به نتایج برخی از مطالعات گذشته (مانند جانسن و رزنیچک^۹؛ ۲۰۰۶ و بوشر^{۱۰} و همکاران، ۲۰۱۰) تاکید می‌کند که کاربران نتایج اصلی و خالصی که بر اساس معیارهای ربط و به طور خودکار به واسطه الگوریتم‌های رتبه‌بندی موتورهای جستجو محاسبه می‌شوند، را ترجیح می‌دهند. نتایج مطالعه علی^{۱۱} و گول^{۱۲} (۲۰۱۶) با عنوان کارایی موتورهای کاوش با استفاده از طبقه‌بندی پرس وجوها در حوزه کتابداری و اطلاع‌رسانی نیز نشان داد که موتور کاوش گوگل به لحاظ مانعیت (میانگین=۱/۱۰) در مقایسه با یاهو (میانگین=۰/۸۸) بهتر است. به لحاظ جامعیت نسبی نیز گوگل (میانگین=۰/۶۸) در مقایسه با یاهو (میانگین=۰/۳۱) از وضعیت بهتری برخوردار است.

^۱Shafi

^۲Rather

^۳SamPATH Kumar

^۴Prakash

^۵Wang

^۶Domachowski

^۷Griesbaum

^۸Heuwing

^۹Resnick

^{۱۰}Buscher

^{۱۱}Ali

^{۱۲}Gul

روش شناسی

پژوهش از نظر هدف کاربردی و به لحاظ روش پیمایشی توصیفی است. محققان با رعایت قالب فرمول جستجو و عبارت جستجو به عنوان شرایط یکسان (عملگر AND و فرمت PDF)، کارآیی موتورهای کاوش وب را بر اساس آنچه در نتایج بازیابی مشاهده می‌شود ارزیابی کرده است. برای اندازه‌گیری متغیرها از فرمول‌های مختلف مربوط به جامعیت نسبی، مانعیت، ضریب F نسبت پوشش، نسبت تازگی، کاوش مورد انتظار و ناکامی استفاده شده است. بافت این مطالعه موتورهای کاوش وب شامل گوگل، بینگ، یاهو، اسک و ای.او.ال. به عنوان موتورهای جستجوی عمومی پرکاربرد بر اساس رتبه‌بندی آلکسا است. ابزار گردآوری داده‌ها در این پژوهش شامل مطالعه کتابخانه‌ای برای تدوین ادبیات پژوهش و چک لیست و مشاهده مستقیم نتایج جست‌وجو از موتورهای کاوش است. برای این منظور بر اساس مقاله 2016 top trends in academic libraries در مجله 'College and Research Libraries' واقع در پایگاه SAGE، حوزه‌های جدید و گرایش‌های آتی علم اطلاعات در قالب ۱۵ عبارت، مشخص گردید. سپس این ۱۵ حوزه جمع‌بندی و جهت تعیین کلیدواژه‌های جست‌وجو در اختیار ۲۰ نفر از متخصصان رشته قرار گرفت تا به لحاظ ارزش قرار گرفتن در فهرست جستجوی این پژوهش نمره وزنی بدهند. سپس ۵ کلیدواژه دارای وزن بیشتر برای جست‌وجو انتخاب شدند. در ادامه هریک از ۵ کلیدواژه به صورت مجزا در موتورهای کاوش پنجگانه مورد جستجو قرار گرفتند. در مرحله بعد از مجموع نتایج بازیابی شده هر موتور کاوش برای هر یک از کلیدواژه‌ها، ۵۰ مدرک ابتدای لیست بازیابی برای ارزیابی معیارهای کارآیی بازیابی در گردآورد محقق ساخته قرار گرفت. برای تشخیص ربط مدارک، تکرار کلیدواژه در عنوان مدرک، چکیده و کلیدواژه‌های آن ملاک قرار گرفت. برای ارزیابی پایایی نتایج بازیابی، از روش بازآزمایی استفاده شد. بدین منظور طی دو مرحله با فاصله ۱۵ روز جستجو و بازیابی مجدداً انجام گردید و همبستگی نتایج دو آزمون مورد آزمون قرار گرفت و در مقدار $R = 89$ مورد تأیید قرار گرفت. برای تجزیه و تحلیل داده‌ها از نرم‌افزار اکسل ۲۰۱۳ استفاده شد.

با این توضیح که، به منظور محاسبه مانعیت موتورهای کاوش پنج‌گانه، از تعداد کل مدارک بازیابی شده، مدارک مرتبط و غیرمرتبط بازیابی شده بر اساس موتور کاوش مشخص شدند. در نهایت فرمول مانعیت نسبی در نرم‌افزار اکسل فرمول‌نویسی و محاسبه مانعیت نسبی انجام گردید. برای بررسی شاخصه جامعیت نسبی تعداد مدارک مرتبط برای هر کلیدواژه از گردآورد ۲۵۰ پیشینه‌ای حاصل از نتایج موتورهای کاوش پنجگانه مشخص شد. سپس بر اساس موتور کاوش بازیابی کننده مرتب گردید تا تعداد مدارک مرتبط بازیابی شده هر موتور جستجو مشخص گردد. در نهایت فرمول جامعیت نسبی در نرم‌افزار اکسل فرمول‌نویسی و محاسبه جامعیت نسبی انجام گردید.

برای محاسبه ضریب کاوش مورد انتظار، هر کدام از کلیدواژه‌ها در موتورهای کاوش موردنظر وارد شدند و مدت زمان بازیابی مدارک برای هر موتور کاوش ثبت شد و در نهایت میانگین مدت زمان بازیابی هر موتور کاوش برای هر کلیدواژه محاسبه گردید. به منظور محاسبه میزان ناکامی، تعداد مدارک مرتبط بازیابی شده هر موتور کاوش از تعداد کل مدارک بازیابی شده (۵۰ مدرک) کسر گردید؛ و برای محاسبه ضریب پوشش تعداد اسناد انحصاری از تعداد کل اسناد مرتبط هر موتور جستجو کسر گردید سپس تعداد اسناد باربط بازیابی شده هر موتور کاوش بر تعداد اسناد آن تقسیم شد. هم‌چنین برای محاسبه نسبت تازگی، مدارکی که تنها توسط یک موتور کاوش بازیابی شده باشند در نظر گرفته شد. اگر مدارکی را دو یا بیش از دو موتور بازیابی کرده باشند جزء مدارک انحصاری هیچ کدام محاسبه نشدند.

یافته‌های پژوهش

با هدف بررسی کارآیی موتورهای کاوش پنج‌گانه به لحاظ شاخصه جامعیت نسبی، داده‌ها با استفاده از فرمول محاسبه جامعیت نسبی بشرح زیر تحلیل گردید که نتایج در جدول ۱ آمده است. نتایج جدول ۱ نشان می‌دهد که وضعیت شاخصه جامعیت نسبی موتور کاوش گوگل دربازایی مدارک مرتبط با کلیدواژه‌های مختلف هم در سطح خرد و هم کلی بهتر از سایر موتورهای کاوش است لذا می‌توان گفت موتورهای کاوش گوگل و اسک در مقایسه با سایر موتورهای کاوش مورد مطالعه توانسته‌اند نسبتاً مدارک مرتبط بیشتری را بازایی نمایند.

بهمی و فرهادپور: بررسی میزان کارایی موتورهای کاوش وب در بازیابی...

جدول ۱ کارایی موتورهای کاوش پنجگانه به لحاظ شاخصه جامعیت نسبی در بازیابی اطلاعات مربوط به حوزه اطلاعات و دانش

میانگین جامعیت نسبی	Integrated Library System				Information Fluency				Professional Development for Librarians				Data Management				Collection Assessment Trends				موتور کاوش
	r	RRDX	PRDR	PRD	r	RRDX	PRDR	PRD	r	RRDX	PRDR	PRD	r	RRDX	PRDR	PRD	r	RRDX ^۲	PRDR ^۳	PRD ^۴	
۰/۲۴	۰/۲	۴۹	۲۳۸	۲۵۰	۰/۲۳	۴۸	۲۰۷	۲۵۰	۰/۲۳	۳۳	۱۴۰	۲۵۰	۰/۲۲	۴۷	۲۱۴	۲۵۰	۰/۳۱۹	۲۹	۹۱	۲۵۰	گوگل
۰/۱۸	۰/۲	۴۹	۲۳۸	۲۵۰	۰/۲	۴۲	۲۰۷	۲۵۰	۰/۱۶	۲۳	۱۴۰	۲۵۰	۰/۱۹۶	۴۲	۲۱۴	۲۵۰	۰/۱۳۲	۱۲	۹۱	۲۵۰	ياهو
۰/۱۷	۰/۱۸	۴۴	۲۳۸	۲۵۰	۰/۱۴	۲۹	۲۰۷	۲۵۰	۰/۲	۲۹	۱۴۰	۲۵۰	۰/۱۸۲	۳۹	۲۱۴	۲۵۰	۰/۱۳۲	۱۲	۹۱	۲۵۰	بینگ
۰/۱۸	۰/۱۹	۴۷	۲۳۸	۲۵۰	۰/۲	۴۲	۲۰۷	۲۵۰	۰/۱۹	۲۷	۱۴۰	۲۵۰	۰/۲	۴۴	۲۱۴	۲۵۰	۰/۱۴۳	۱۳	۹۱	۲۵۰	ای.اوال
۰/۲۲	۰/۲	۴۹	۲۳۸	۲۵۰	۰/۲۲	۴۶	۲۰۷	۲۵۰	۰/۲	۲۸	۱۴۰	۲۵۰	۰/۱۹۱	۴۱	۲۱۴	۲۵۰	۰/۲۷۵	۲۵	۹۱	۲۵۰	اسک

۱ Pool of Retrieved Documents (گردآورد مدارک بازیابی شده)

۲ Pool of Relevant Retrieved Document (گردآورد مدارک مرتبط بازیابی شده)

۳ Relevant Retrieved Documents of X Search Engine (مدارک مرتبط موتور کاوش X)

۴ جامعیت نسبی

با هدف بررسی کارآیی موتورهای کاوش پنج گانه به لحاظ شاخصه مانعیت، داده‌ها با استفاده از فرمول محاسبه مانعیت تحلیل گردید که نتایج در جدول شماره ۲ آمده است.

$$P = \frac{|Ra|}{|A|}$$

جدول ۲ کارایی موتورهای کاوش پنجگانه به لحاظ شاخصه مانعیت در بازیابی اطلاعات مربوط به حوزه اطلاعات و دانش

موتور کاوش	Collection Assessment Trends			Data Management			Professional Development for Librarians			Information Fluency			Integrated Library System		
	P ^۱	RRDX ^۲	RDX ^۳	P	RRDX	RDX	P	RRDX	RDX	P	RRDX	RDX	P	RRDX	RDX
گوگل	۰/۵۸	۲۹	۵۰	۰/۹۴	۴۷	۵۰	۰/۶۶	۳۳	۵۰	۰/۹۶	۴۸	۵۰	۰/۹۸	۴۹	۵۰
ياهو	۰/۲۴	۱۲	۵۰	۰/۸۴	۴۲	۵۰	۰/۴۶	۲۳	۵۰	۰/۸۴	۴۲	۵۰	۰/۹۸	۴۹	۵۰
بينگ	۰/۲۴	۱۲	۵۰	۰/۷۸	۳۹	۵۰	۰/۵۸	۲۹	۵۰	۰/۵۸	۲۹	۵۰	۰/۸۸	۴۴	۵۰
ای.اوال	۰/۲۶	۱۳	۵۰	۰/۸۸	۴۴	۵۰	۰/۵۴	۲۷	۵۰	۰/۸۴	۴۲	۵۰	۰/۹۴	۴۷	۵۰
اسک	۰/۵	۲۵	۵۰	۰/۸۲	۴۱	۵۰	۰/۵۶	۲۸	۵۰	۰/۹۲	۴۶	۵۰	۰/۹۸	۴۹	۵۰

براساس نتایج جدول ۲ به لحاظ شاخصه مانعیت چه در سطح تک تک کلیدواژه‌های جستجو شده و چه در سطح میانگین کل وضعیت موتور کاوش گوگل بهتر از سایر موتورهای کاوش است و موتورهای کاوش یاهو و ای.اوال در رتبه پایین قرار دارند. لذا می‌توان گفت موتور کاوش گوگل (میانگین = ۰/۸۲۴) در مقایسه با سایر موتورهای کاوش مورد نظر موفقیت بیشتری در نادیده گرفتن مدارک نامرتبط یا به عبارتی دقت جستجو داشته است.

با هدف بررسی کارآیی موتورهای کاوش پنج گانه به لحاظ شاخصه ضریب F (میانگین توافقی)، داده‌ها با استفاده از فرمول میانگین توافقی محاسبه گردید که نتایج در جدول شماره ۳ آمده است.

$$F(j) = \frac{2}{\frac{1}{r(j)} + \frac{1}{p(j)}}$$

طبق نتایج جدول ۳ به لحاظ شاخصه میانگین توافقی موتور کاوش گوگل هم در سطح کلیدواژه‌ها به صورت انفرادی و میانگین کل نمره بالاتری کسب کرده است و موتورهای اسک، ای.اوال،، یاهو و بینگ در رتبه‌های بعدی قرار گرفتند. با توجه به مقدار میانگین توافقی کلی، می‌توان گفت موتور کاوش گوگل در مقایسه با سایر موتورهای کاوش توانسته است ترکیب بهتری میان جامعیت و مانعیت برقرار کند.

۱ مدارک بازیابی شده توسط موتور کاوش X (RDX=Retrieved Documents of X Search Engine)
 ۲ مدارک مرتبط بازیابی شده توسط موتور کاوش X (Relevant Retrieved Documents of X Search Engine)
 ۳ مانعیت

جدول ۳ کارآیی موتورهای کاوش پنج‌گانه به لحاظ شاخصه ضریب F در بازیابی اطلاعات مربوط به حوزه اطلاعات و دانش

میانگین توافقی کلی	Integrated Library System			Information Fluency			Professional Development for Librarians			Data Management			Collection Assessment Trends			موتور کاوش
	F	P(j)	R(j)	F	P(j)	R(j)	F	P(j)	R(j)	F	P(j)	R(j)	F	P(j)	R(j)	
۰/۳۶	۰/۳۳	۱/۰۲	۵	۰/۳۷	۱/۰۴	۴/۳۴	۰/۳۴	۱/۵۱	۴/۳۴	۰/۳۵	۱/۰۶	۴/۵۴	۰/۴۱	۱/۷۲	۳/۱۳	گوگل
۰/۲۷۲	۰/۳۳	۱/۰۲	۵	۰/۳۲	۱/۱۹	۵	۰/۲۳	۲/۱۷	۶/۲۵	۰/۳۱	۱/۱۹	۵/۱۰	۰/۱۷	۴/۱۶	۵/۵۷	ياهو
۰/۲۵۲	۰/۲۹	۱/۱۳	۵/۵۵	۰/۲۲	۱/۷۲	۷/۱۴	۰/۲۹	۱/۷۲	۵	۰/۲۹	۱/۲۸	۵/۴۹	۰/۱۷	۴/۱۶	۷/۵۷	بينگ
۰/۲۸۲	۰/۳۱	۱/۰۶	۵/۲۶	۰/۳۲	۱/۱۹	۵	۰/۲۸	۱/۸۵	۵/۲۶	۰/۳۲	۱/۱۳	۵	۰/۱۸	۳/۸۴	۶/۹۹	ای.اوا.ل
۰/۳۲۴	۰/۳۳	۱/۰۲	۵	۰/۳۵	۱/۰۸	۴/۵۴	۰/۲۹	۱/۷۸	۵	۰/۳۰	۱/۲۱	۵/۲۳	۰/۳۵	۲	۳/۶۳	اسک

با هدف بررسی کارآیی موتورهای کاوش پنج‌گانه به لحاظ شاخصه نسبت پوشش، داده‌ها با استفاده از فرمول محاسبه نسبت پوشش تحلیل گردید که نتایج در جدول شماره ۴ آمده است.

$$C_r = \frac{|RK|}{|U|}$$

جدول ۴ کارآیی موتورهای کاوش پنج‌گانه به لحاظ شاخصه نسبت پوشش در بازیابی اطلاعات مربوط به حوزه اطلاعات و دانش

میانگین توافقی کلی	Integrated Library System			Information Fluency			Professional Development for Librarians			Data Management			Collection Assessment Trends			موتور کاوش
	Cr	RK	U	Cr	RK	U	Cr	RK	U	Cr	RK	U	Cr	RK	U	
۰/۵۱۸	۰/۷۱	۳۵	۴۹	۰/۶۶	۳۲	۴۸	۰/۶۰	۲۰	۳۳	۰/۳۸	۱۸	۴۷	۰/۵۵	۱۶	۲۹	گوگل
۰/۷۱۴	۰/۸۷	۴۳	۴۹	۰/۷۳	۳۱	۴۲	۰/۵۲	۱۲	۲۳	۰/۵۴	۲۳	۴۲	۰/۹۱	۱۱	۱۲	ياهو
۰/۶۲۸	۰/۷۰	۳۱	۴۴	۰/۶۲	۱۸	۲۹	۰/۸۶	۲۵	۲۹	۰/۳۰	۱۲	۳۹	۰/۶۶	۸	۱۲	بينگ
۰/۶۹۸	۰/۸۰	۳۸	۴۷	۰/۶۹	۲۹	۴۲	۰/۶۶	۱۸	۲۷	۰/۵	۲۲	۴۴	۰/۸۴	۱۱	۱۳	ای.اوا.ل
۰/۵۶	۰/۷۱	۳۵	۴۹	۰/۵۲	۲۴	۴۶	۰/۵۷	۱۶	۲۸	۰/۳۶	۱۵	۴۱	۰/۶۴	۱۶	۲۵	اسک

ه لحاظ شاخصه نسبت پوشش داده‌های جدول ۴ نشان می‌دهد که موتور کاوش ياهو نمره بالاتری کسب کرده است و سپس موتورهای کاوش ای.اوا.ل، بینگ، اسک و گوگل در رتبه‌های بعدی قرار دارند. لذا می‌توان گفت موتور جست‌وجوی ياهو در مقایسه با سایر موتورهای کاوش مورد نظر توانسته است نسبتاً مدارک با ربط شناخته شده بیشتری را بازیابی نماید که این مدارک در بازیابی سایر موتورهای کاوش نیز حضور داشته‌اند.

با هدف بررسی کارآیی موتورهای کاوش پنج‌گانه به لحاظ شاخصه نسبت تازگی، داده‌ها با استفاده از فرمول محاسبه نسبت تازگی تحلیل گردید که نتایج در جدول شماره ۵ آمده است.

$$N_r = \frac{|Ru|}{|Ru| + |Rk|}$$

جدول ۵ کارآیی موتورهای کاوش پنج‌گانه به لحاظ شاخصه نسبت تازگی در بازیابی اطلاعات مربوط به حوزه اطلاعات و دانش

میانگین نسبت تازگی کلی	Integrated Library System				Information Fluency				Professional Development for Librarians				Data Management				Collection Assessment Trends				موتور کاوش
	Novelty Rate	RK	RU	RD	Novelty Rate	RK	RU	RD	Novelty Rate	RK	RU	RD	Novelty Rate	RK	RU	RD	Novelty Rate	RK	RU	RD	
۰/۴۱	۰/۲۸	۳۵	۱۴	۴۹	۰/۳۳	۳۲	۱۶	۴۸	۰/۳۹	۲۰	۱۳	۳۳	۰/۶۱	۱۸	۲۹	۴۷	۰/۴۴	۱۶	۱۳	۲۹	گوگل
۰/۲۷	۰/۱۲	۴۳	۶	۴۹	۰/۲۶	۳۱	۱۱	۴۲	۰/۴۷	۱۲	۱۱	۲۳	۰/۴۵	۲۳	۱۹	۴۲	۰/۰۸	۱۱	۱	۱۲	ياهو
۰/۳۶	۰/۲۹	۳۱	۱۳	۴۴	۰/۳۷	۱۸	۱۱	۲۹	۰/۱۳	۲۵	۴	۲۹	۰/۶۹	۱۲	۲۷	۳۹	۰/۳۳	۸	۴	۱۲	بينگ
۰/۲۹	۰/۱۹	۳۸	۹	۴۷	۰/۳۰	۲۹	۱۳	۴۲	۰/۳۳	۱۸	۹	۲۷	۰/۵	۲۲	۲۲	۴۴	۰/۱۵	۱۱	۲	۱۳	ای.اوال
۰/۴۳	۰/۲۸	۳۵	۱۴	۴۹	۰/۴۷	۲۴	۲۲	۴۶	۰/۴۲	۱۶	۱۲	۲۸	۰/۶۳	۱۵	۲۶	۴۱	۰/۳۶	۱۶	۹	۲۵	اسک

نتایج جدول ۵ نشان می‌دهد که به لحاظ نسبت تازگی (درصد بازیابی مدارک مرتبط جدیدی که در نتایج سایر موتورهای کاوش مشاهده نشده است) به طور کلی موتور کاوش اسک و سپس گوگل نمره بالاتری را کسب کرده‌اند. از این حیث موتور کاوش یاهو نمره پایینی کسب کرده است. لذا می‌توان گفت موتور کاوش اسک در مقایسه با سایر موتورهای مورد مطالعه توانایی نسبتاً بیشتری در بازیابی اسناد مرتبط ناشناخته و جدید دارد.

با هدف بررسی کارآیی موتورهای کاوش پنج‌گانه به لحاظ شاخصه ضریب کاوش مورد انتظار، سوال ۶ طرح و داده‌ها با استفاده از فرمول محاسبه ضریب کاوش مورد انتظار تحلیل گردید که نتایج در جدول شماره ۶ آمده است.

جدول ۶ کارآیی موتورهای کاوش پنج‌گانه به لحاظ شاخصه ضریب کاوش مورد انتظار در بازیابی اطلاعات مربوط به حوزه اطلاعات و دانش

ضریب کاوش مورد انتظار (ثابته)	Integrated Library System	Information Fluency	Professional Development for Librarians	Data Management	Collection Assessment Trends	موتور کاوش
۰/۶۱	۰/۶۵	۰/۵۲	۰/۶۴	۰/۷۶	۰/۵۲	گوگل
۰/۶۲	۰/۴۸	۰/۷۵	۰/۷۴	۰/۶۱	۰/۵۵	ياهو
۰/۷۰	۰/۶۰	۰/۸۲	۰/۹۱	۰/۵۷	۰/۶۲	بينگ
۰/۶۶	۰/۶۷	۰/۸۱	۰/۶۹	۰/۴۴	۰/۷۲	ای.اوال
۰/۶۰	۰/۶۳	۰/۶۵	۰/۷۱	۰/۵۶	۰/۴۸	اسک

به لحاظ ضریب کاوش مورد انتظار یا طول کاوش نتایج جدول ۶ نشان می‌دهد که موتورهای جستجوی اسک (میانگین = ۰/۶۰) و گوگل (۰/۶۱) در مقایسه با سایر موتورهای کاوش عملکرد بهتری داشته‌اند. لذا می‌توان گفت موتور کاوش اسک و گوگل در مقایسه با سایر موتورهای مورد مطالعه سرعت بیشتری در بازیابی مدارک مربوطه داشته است.

با هدف بررسی کارآیی موتورهای کاوش پنج‌گانه به لحاظ شاخصه ناکامی، داده‌ها با استفاده از فرمول محاسبه ناکامی تحلیل گردید که نتایج در جدول شماره ۷ آمده است.

۱ مدارک مرتبطی که یک موتور کاوش از مجموع کل مدارک مرتبط بازیابی شده به خود اختصاص داده است.

۲ مدارک مرتبطی که صرفاً در نتایج بازیابی یک موتور کاوش (انحصاری) دیده شده است.

۳ مدارک مرتبطی که به طور مشترک در نتایج سایر موتورها (حداقل یک موتور کاوش دیگر) هم دیده شده است.

جدول ۷ کارایی موتورهای کاوش پنج‌گانه به لحاظ شاخصه ناکامی در بازیابی اطلاعات مربوط به حوزه اطلاعات و دانش

میانگین ناکامی کلی	Integrated Library System			Information Fluency			Professional Development for Librarians			Data Management			Collection Assessment Trends			موتور کاوش
	Fr	RRDX	RDX	Fr	RRDX	RDX	Fr	RRDX	RDX	Fr	RRDX	RDX	Fr	RRDX	RDX	
	۸/۸	۱	۴۹	۵۰	۲	۴۸	۵۰	۱۷	۲۳	۵۰	۲	۴۷	۵۰	۲۱	۲۹	
۱۶/۴	۱	۴۹	۵۰	۸	۴۲	۵۰	۲۷	۲۳	۵۰	۸	۴۲	۵۰	۳۸	۱۲	۵۰	ياهو
۱۹/۴	۶	۴۴	۵۰	۲۱	۲۹	۵۰	۲۱	۲۹	۵۰	۱۱	۳۹	۵۰	۳۸	۱۲	۵۰	بینگ
۱۵/۴	۳	۴۷	۵۰	۸	۴۲	۵۰	۲۳	۲۷	۵۰	۶	۴۴	۵۰	۳۷	۱۳	۵۰	ای.اوال
۱۲/۲	۱	۴۹	۵۰	۴	۴۶	۵۰	۲۲	۲۸	۵۰	۹	۴۱	۵۰	۲۵	۲۵	۵۰	اسک

به لحاظ شاخصه ناکامی (توجه به میزان اسناد بی‌ربط بازیابی شده) نتایج جدول ۷ نشان می‌دهد که موتور کاوش گوگل (میانگین=۸/۸) نمره بهتری را در مقایسه با سایر موتورهای کاوش کسب کرده است و موتور کاوش بینگ در رتبه آخر قرار گرفته است. لذا می‌توان گفت موتور کاوش گوگل مدارک نامرتب کمتري بازیابی نموده است.

نتیجه‌گیری و بحث

در این مطالعه شاخصه‌های جامعیت نسبی (نسبت بین تعداد اسناد با ربط بازیابی شده توسط سامانه و تعداد اسنادی که کاربر انتظار داشته بتواند بازیابی کند)، مانعیت (بخشی از اسناد بازیابی شده از مجموعه (A) که با ربط هستند و یا به عبارتی توانایی نظام در نادیده گرفتن مدارک نامرتب)، ضریب F یا میانگین توافقی (سنجه واحدی که جامعیت و مانعیت را ترکیب و هدف آن یافتن بهترین ترکیب ممکن بین جامعیت و مانعیت است)، نسبت پوشش (کسری از اسناد با ربط شناخته شده برای کاربر تعریف می‌شود که عملاً در نتیجه یک جستجو بازیابی شده است)، ضریب کاوش مورد انتظار (طول کاوش مورد انتظار که برای کار با مجموعه‌های اسنادی با نظم ضعیف به کار می‌رود)، ناکامی (توجه به میزان اسناد بی‌ربط بازیابی شده) (بیزایتس و ریرو، ۱۳۸۵)، و نسبت تازگی و نوظهوری (بخشی از مدارک جدید که برای اولین بار مورد توجه کاوشگر اطلاعات قرار گرفته است) (اشرفی‌ریزی و کاظم‌پور، ۱۳۸۶) موتورهای کاوش پنجگانه گوگل، یاهو، بینگ، ای.اوال. و اسک بر اساس رتبه‌بندی آلکسا و برای جستجوهای مبتنی بر کلیدواژه‌های حوزه اطلاعات و دانش مورد مطالعه قرار گرفت.

نتایج نشان داد که به لحاظ جامعیت نسبی، مانعیت، ضریب F یا میانگین توافقی و ضریب ناکامی موتور جستجوی گوگل، به لحاظ شاخصه‌های ضریب کاوش مورد انتظار و نسبت تازگی موتور کاوش اسک، و به لحاظ نسبت پوشش موتور جستجوی یاهو عملکرد بهتری در مقایسه با سایر موتورهای مورد مطالعه داشته‌اند. با توجه به این که در مطالعات گذشته موتورهای کاوش بررسی شده، حوزه علمی و شاخصه‌های کارایی متفاوت بوده‌اند و عمدتاً بر جامعیت نسبی و مانعیت تاکید شده است، لذا امکان مقایسه تمام یافته‌ها فراهم نیست؛ لیکن یافته‌های این مطالعه بر اساس شاخصه جامعیت نسبی با یافته‌های مطالعات شافی و رادر (۲۰۰۵) و علی و گول (۲۰۱۶)؛ بر اساس مانعیت با یافته‌های جانسن و مولینا (۲۰۰۶)، کومار و پراکاش (۲۰۰۹)، ونگ و همکاران (۲۰۱۲) و علی و گول (۲۰۱۶)؛ و بر اساس شاخصه نسبت پوشش با یافته‌های محمداسماعیل و منصورکیایی (۱۳۹۰)، اسفندیاری مقدم (۱۳۹۱) و جانسن و مولینا (۲۰۰۶) همسو است.

نتیجه این که، کارایی موتورهای کاوش علی‌رغم نمره‌ای که کسب کردند و در این مطالعه به رتبه‌بندی آنها منجر شد، اما با وضعیت آرمانی فاصله زیادی دارند. این مساله نشان می‌دهد چالش‌های مربوط به ارزیابی کارایی بازیابی اطلاعات، علی‌رغم

وجود موتورهای کاوش مختلف، بکارگیری استراتژی‌های گوناگون، الگوریتم‌های مختلف رتبه‌بندی و شیوه‌های نمایه‌سازی اسناد وبی هم‌چنان محل چالش و نیازمند مطالعات بیشتر است. نکته دیگر این که، مقایسه نتایج این مطالعه با بخشی از یافته‌های مطالعات گذشته بیانگر این واقعیت است که هیچ‌کدام از موتورهای کاوش به تنهایی شاخص‌های کارایی را برآورده نمی‌کنند. لذا کاربران با توجه به شاخصه‌های مختلف بایستی پرس‌وجوی خود را به موتورهای کاوش مختلف بسپارند. ضمن این که طرفی نیاز به طراحی موتورهای کاوش تخصصی با عنایت به گستره، تنوع و ارتباط واژگانی برای حوزه‌های علمی مختلف ضروری است؛ از طرف دیگر، به نظر می‌رسد زمان آن رسیده باشد که استفاده از الگوهای مختلف مانند جستجوهای بصری و یا جستجو و بازیابی براساس نمایه‌سازی وزنی در رابط جستجوی موتورهای کاوش عمومی و تخصصی بیش از پیش مورد توجه قرار گیرد. نهایتاً با توجه به یافته‌های این مطالعه پیشنهاد می‌شود کاربرانی که برای نیاز اطلاعاتی خود در حوزه علم اطلاعات و دانش‌شناسی از موتورهای کاوش عمومی استفاده می‌کنند:

- با توجه به این که عمده کاربرانی که از موتورهای کاوش عمومی استفاده می‌کنند، همه مدارک بازیابی شده را بازبینی و ارزیابی نمی‌کنند و فقط به مدارک محدود ابتدای لیست (یا حداکثر چند صفحه ابتدای فهرست بازیابی) توجه دارند؛ در صورتی که به دنبال بازیابی مدارک مرتبط بیشتری هستند (و البته در کنار این مساله بازیابی مدارک نامرتب نیز قابل تحمل است) از موتور کاوش گوگل استفاده نمایند. به همین ترتیب اگر دقت بازیابی (قابلیت نادیده گرفتن مدارک غیرمرتبط) و یا ناکامی کمتر نیز ملاک مهمی است، موتور کاوش گوگل عملکرد بهتری دارد.
- در صورتی که کاربران به دنبال ترکیب مناسبی از جامعیت و مانعیت هستند، استفاده از موتورهای کاوش گوگل و اسک به خاطر نمره کارایی آنها در میانگین توافقی توصیه می‌شود.
- در صورتی که مدت زمان بازیابی ملاک مدنظر باشد، استفاده از موتورهای کاوش اسک و گوگل پیشنهاد می‌شود.
- در صورتی که کاربران خواهند با جستجوی یک موتور کاوش به نتایجی دست یابد که بخشی از نتایج سایر موتورهای کاوش را پوشش دهد، استفاده از موتور کاوش یاهو پیشنهاد می‌شود.
- در صورتی که کاربر به دنبال بازیابی مدارک تازه (ناشناخته) است، استفاده از موتور کاوش اسک توصیه می‌شود.

منابع

- اسفندیاری مقدم، علیرضا. (۱۳۹۱). میزان هم‌پوشانی نتایج بازیابی شده کلیدواژه‌های تخصصی پزشکی در موتورهای کاوش عمومی وب. *مدیریت اطلاعات سلامت*، ۲۴، ۲۰۳-۲۱۴.
- اشرفی ریزی، حسن، کاظم پور، زهرا. (۱۳۸۶). نقش و کاربرد تفکر انتقادی در ارزیابی منابع اینترنتی. *اطلاع‌شناسی*، ۱۷ و ۱۸، ۱۱۹-۱۳۲.
- بیزایتس، ریکاردو، ربرو، برتیه. (۱۳۸۵). قلمروهای نو در بازیابی اطلاعات (ترجمه علی حسین قاسمی). تهران: چاپار.
- چشمه‌سهرابی، مظفر. (۱۳۷۸). *تأثیر استفاده از اصطلاحنامه در بانکهای اطلاعاتی کتابشناختی بر میزان جامعیت، مانعیت و مدت زمان جستجوی اطلاعات بازیابی شده*. پایان‌نامه کارشناسی ارشد، استاد راهنما عباس حری، دانشگاه تربیت مدرس: دانشکده علوم تربیتی و روانشناسی.

- شاگری، صدیقه. (۱۳۸۷). میزان جامعیت و مانعیت ابزارهای کاوش فارسی اینترنت در بازیابی اطلاعات در حوزه کتابداری و اطلاع‌رسانی. فصلنامه کتاب، ۷۳، ۱۷۷-۲۰۰.
- صراطی شیرازی، منصوره. (۱۳۸۸). مقایسه میزان دقت موتورهای کاوش عمومی و تخصصی پزشکی در بازیابی مدارک مربوط به بیماری‌های کودکان. فصلنامه کتاب، ۷۷، ۷۷-۹۴.
- کوشا، کیوان. (۱۳۸۱). ابزارهای کاوش اینترنت: اصول، مهارت‌ها و امکانات جستجو در وب، تهران: نشر کتابدار.
- محمداسماعیل، صدیقه، منصورکیایی، ربابه. (۱۳۹۰). مقایسه موتورها و ابرموتورهای کاوش عمومی در بازیابی اطلاعات علم فیزیک و میزان هم‌پوشانی آن‌ها. مطالعات ملی کتابداری و سازماندهی اطلاعات، شماره ۸۷، ۱۳۰-۱۴۰.
- Ali, S. & Gul, S. (2016). Search engine effectiveness using query classification: A study. *Online Information Review*, 40(4), 515-528. Doi:<https://doi.org/10.1108/OIR-07-2015-0243>.
- Ali, T., Jhandir, Z., Lee, I., On, B-W. & Choi, G. S. (2017). Evaluating retrieval effectiveness by sustainable rank list. *Sustainability*, 9, 1203, 1-20. Doi:10.3390/su9071203.
- Azimzadeh, M., Badie, R. & Esnaashari, M. M. (2016). A review on web search engines' automatic evaluation methods and how to select the evaluation method. 2nd International Conference on Web Research, ICWR, 27-28 April, Tehran, pp78-83.
- Croft, W. B., Metzler, D. & Strohman, T. (2015). *Search Engines: Information Retrieval in Practice*. London: Pearson Education, Inc, 518 pages.
- Domachowski, A., Griesbaum, J. & Heuwing, B. (2015). Perception and effectiveness of search advertising on smartphones. *Proceedings of the Association for Information Science and Technology*, 53(1), 1-10. Doi: 10.1002/pr2.2016.14505301074.
- Evans, M. P. (2007). Analysing Google rankings through search engine optimization data. *Internet Research*, 17(1), 21-37. Doi:<http://dx.doi.org/10.1108/10662240710730470>
- Gross, A. M. (2015). Information retrieval in Arabic: An evaluation of three multilingual search engines on their capabilities in dealing with Arabic search queries. *International Journal of Information Technology and Business Management*, 4(1), 1-21.
- Jansen, B. J. & Molina, P. R. (2006). The effectiveness of web search engines for retrieving relevant ecommerce links. *Information Processing and Management*, 42, 1075-1098. doi: 10.1016/j.ipm.2005.09.003.
- Katumba, S. & Coetzee, S. (2017). Employing search engine optimization (SEO) techniques for improving the discovery of geospatial resources on the web. *International Journal of Geo-Information*, 6, 284, 1-20. doi:10.3390/ijgi6090284
- Khorsheed, K. O., Madbouly, M. & Guirguis, S. K. (2015). Search engine optimization using data mining approach. *International Journal of Computer Engineering and Applications*, 9(6-1), 184-200.
- Kulkarni, A. (2013). Efficient and effective large-scale search. PhD. Thesis, Advisor: Jamie Callan. The university of Melbourne, Language and Information Technologies Department. 167 pages. Retrieved 06/07/2017 from: www.lti.cs.cmu.edu.
- Levene, M. (2010). *An Introduction to Search Engines and Web Navigation*. [2nd Edition]. New Jersey: John Wiley & Sons. 463 pages.
- Mivule, K. (2017). Web search query privacy, an end-user perspective. *Journal of Information Security*, 8, 56-74. Retrieved 06/07/2017 from: <http://dx.doi.org/10.4236/jis.2017.81005>

- Sahu, S. K., Mahapatra, D. P. & Balabantray, R. C. (2016). Comparative study of search engines in context of features and semantics. *Journal of Theoretical and Applied Information Technology*, 88(2), 210-218.
- Sampath kumar, B. T. & Prakash, J. N. (2009). Precision and relative recall of search engines: A comparative study of Google and Yahoo. *Singapore Journal of Library & Information Management*, 38, 124-137.
- Shafi, S. M., & Rather, R. A. (2005). Precision and recall of five search engines for retrieval of scholarly information in the field of biotechnology. *Webology*, 2 (2), Article 12. Available at: <http://www.webology.org/2005/v2n2/a12.html>.
- Sibuyi, T. & Dehinbo, J. O. (2016). Optimization and effectiveness of search engine results. Proceedings of the World Congress on Engineering and Computer Science, WCECS, 19-21 October, San Francisco, USA. Vol. 1.
- Wang, L., Wang, J., Wang, M., Li, Y., Liang, Y., & Xu, D. (2012). Using internet search engines to obtain medical information: A comparative study. *Journal of Medical Internet Research*, 14(3), e74.
<http://doi.org/10.2196/jmir.1943>
- Wu S., Li J. (2004). Effectiveness evaluation and comparison of web search engines and Meta-search engines. In: Li Q., Wang G., Feng L. (Eds.) *Advances in Web-Age Information Management. WAIM 2004. Lecture Notes in Computer Science*, Vol. 3129. Springer, Berlin, Heidelberg. DOI: https://doi.org/10.1007/978-3-540-27772-9_31.
- Xie, M., Wang, H. & Goh, T. N. (1998). Quality dimensions of search engines. *Journal of Information Science*, 24(5), 365-372. <https://doi.org/10.1177/016555159802400509>.



COPYRIGHTS

©2020 by the authors. Licensee SCU, Ahvaz, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>)