جلد هفتم، شماره دوم (تابستان ۱۳۹۹): ۱۵-۱
http://hii.khu.ac.ir
مقاله اصیل پژوهشی

مطالعه مروری سنجه‌های رتبه‌بندی و غیر رتبه‌بندی تعیین کارآمدی موتورهای کاوش

اعظم صنعت‌جو: دانشیار گروه علم اطلاعات و دانش شناسی، دانشگاه فردوسی مشهد، مشهد، ایران
***مهدی زینالی‌تازه‌کندی:** دانشجوی دکتری علم اطلاعات و دانش شناسی، دانشگاه فردوسی مشهد، مشهد، ایران (نویسنده مسئول)ma.zeynali@mail.um.ac.ir

چکیده

دریافت: ۱۳۹۸/۱۱/۲۰
پذیرش: ۱۳۹۹/۰۲/۰۶

هدف: هدف پژوهش حاضر، تحلیل سنجه‌های مهم و جدید ارزیابی برای به‌کارگیری در موتورهای کاوش به‌منظور ارائه نتایج دقیق است.
روش: مقاله حاضر، مقاله‌ای مروری است و برای گردآوری مقاله‌ها کلیدواژه‌های سنجه‌های ارزیابی، ارزیابی موتورهای کاوش، ارزیابی نظام‌های بازیابی اطلاعات، سنجه‌های ارزیابی ربط در گوگل اسکالر و پایگاه‌های اطلاعاتی مگ ایران و سید جستجو و مقاله‌های مرتبط تهیه شد. در این مطالعه از رویکرد توصیفی- تحلیلی استفاده شد تا سنجه‌های مهم موتورهای کاوش مرور شود.
یافته‌ها: با مرور سنجه‌های مختلف ارزیابی موتورهای کاوش، سه گروه از سنجه‌ها شناسایی شد. می‌توان سنجه‌های دقت، بازیافت را در گروه غیر رتبه‌بندی قرار دارد که متأثر از رویکرد نظام‌گرایی است. از طرف دیگر، گروه رتبه‌بندی شامل سنجه‌های متوسط فاصله، سود تجمعی تعدیل‌یافته نرمال، سنجه کارآمدی رتبه‌بندی و بی پرف است که رویکرد کاربرگرایی در مطرح‌شدن این سنجه‌ها مؤثر بوده است. اگرچه سنجه جامعیت با هدف در نظر گرفتن کاربر ارائه شده است؛ اما به نظر می‌رسد که در عمل به‌صورت کامل به هدف خود نرسیده است. افزون بر این، همان‌گونه که در پژوهش‌های بازیابی اطلاعات، رویکرد سومی نیز مطرح شده است که بر تعامل و یگانش دو رویکرد نظام‌گرا و کاربرگرا اشاره دارد، در سنجه‌های ارزیابی موتورهای کاوش نیز، دو سنجه اتحاد جاکارد و اتحاد کسینوسی برگرفته از رویکرد سوم هست.
نتیجه‌گیری: برای تعیین موتور کاوش کارآمد پیشنهاد می‌شود که پژوهشگران هنگام ارزیابی نظام‌های بازیابی اطلاعات در پژوهش‌های خود، سنجه‌هایی را از هر سه گروه یاد شده انتخاب نمایند.

کلیدواژه‌ها: نظام بازیابی اطلاعات، موتور کاوش، ارزیابی ربط، سنجه‌های ارزیابی، سنجه‌های کارآمدی

تعارض منافع: گزارش نشده است.
منبع حمایت کننده: حامی مالی نداشته است.

شیوه استناد به این مقاله

APA: Sanatjoo, A., zeynali Tazehkandi, A., (2020) Review of ranked-based and unranked-based metrics for determining the effectiveness of search engines. *Human Information Interaction*, 7(2);1-15. (Persian)

Vancouver: Sanatjoo A, zeynali Tazehkandi A. Review of ranked-based and unranked-based metrics for determining the effectiveness of search engines. *Human Information Interaction*. 2020;7(2):1-15. (Persian)

انتشار مجله تعامل انسان و اطلاعات با حمایت مالی دانشگاه خوارزمی انجام می‌شود .
انتشار این مقاله به صورت دسترسی آزاد مطابق با CC BY-NC-SA 3.0 صورت گرفته است.

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

Human Information Interaction. Vol. 7, No. 2 (Summer 2020):1-15.
http://hii.khu.ac.ir
Original Article

Review of Ranked and Unranked-based Metrics for Determining the Effectiveness of Search Engines

Azam Sanatjoo: Associative professor in Knowledge and information science, Department of Knowledge and information science, Ferdowsi University of Mashhad, Mashhad, Iran.
***Mahdi zeynali Tazehkandi:** PhD candidate in Knowledge and information science, Department of Knowledge and information science, Ferdowsi University of Mashhad, Mashhad, Iran. (Corresponding author) ma.zeynali@mail.um.ac.ir

Received: 09/02/2020 Accepted: 25/04/2020

Abstract

Purpose: There are several metrics for evaluating search engines. Though, many researchers have proposed new metrics in recent years. Familiarity with new metrics is essential. So, the purpose is to provide an analysis of important and new metrics to evaluate search engines.

Methodology: This review article critically studied the efficiency of metrics of evaluation. So, "evaluation metrics," "evaluation measure," "search engine evaluation," "information retrieval system evaluation," "relevance evaluation measure" and "relevance evaluation metrics" were investigated in "MagIran" "Sid" and Google Scholar search engines. Articles gathered to inspect and analyse existing approaches in evaluation of information retrieval systems. Descriptive-analytical approach used to review the search engine assessment metrics.

Findings: Theoretical and philosophical foundations determine research methods and techniques. There are two well-known "system-oriented" and "user-oriented" approaches to evaluating information retrieval systems. So, researchers such as Sirotkin (2013) and Bama, Ahmed, & Saravanan (2015) group the precision and recall metrics in a system-oriented approach. They also believe that Average Distance, normalized discounted cumulative gain, Rank Eff and B pref are rooted in the user-oriented approach. Nowkarizi and Zeynali Tazehkandi (2019) introduced comprehensiveness metric instead of Recall metric. They argue that their metric is rooted in a user-oriented approach, while the goal is not fully met. On the other hand, Hjørland(2010) emphasizes that we need a third approach to eliminate this dichotomy. In this regard, researchers such as Borlund, Ingwersen (1998), Borlund (2003), Thornley, Gibb (2007) have mentioned a third approach for evaluating information retrieval systems that refer to interact and compose two mentioned approaches. Incidentally, Borlund, Ingwersen(1998) proposed a Jaccard Association and Cosine Association measures to evaluate information retrieval systems. It seems that these two metrics have failed to compose the system-oriented and user-oriented approaches completely, and need further investigation.

Conclusion: Search engines involve different components including: Crawler, Indexer, Query Processor, Retrieval Software, and Ranker. Scholars wish to apply the most efficient search engines for retrieving required information resources. Each metrics measures a specific component, to measure all, it is suggested to select metrics from all three mentioned groups in their search.

Keywords: Information Retrieval System, Search Engine, Relevance Evaluation, Assessment Measures, Effectiveness Metric.

Conflicts of Interest: None
Funding: None.

How to cite this article

APA: Sanatjoo, A., zeynali Tazehkandi, A., (2020) Review of ranked-based and unranked-based metrics for determining the effectiveness of search engines. *Human Information Interaction*, 7(2);1-15. (Persian)

Vancouver: Sanatjoo A, zeynali Tazehkandi A. Review of ranked-based and unranked-based metrics for determining the effectiveness of search engines. *Human Information Interaction*. 2020;7(2):1-15. (Persian)

The journal of *Human Information Interaction* is supported by Kharazmi University, Tehran, Iran.
This work is published under CC BY-NC-SA 3.0 licence.

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

مقدمه

افراد مختلف در سرتاسر جهان با نیازهای اطلاعاتی مختلف از موتورهای کاوش برای رسیدن به اطلاعات موردنیازشان استفاده می‌کنند که هرکدام از موتورهای کاوش، قابلیت‌ها و توانایی‌های متفاوتی برای بازیابی اطلاعات مرتبط دارند. بر اساس سایت جامع سیاهه موتورهای کاوش[1] (۲۰۱۷)، بیش از هزار موتور کاوش عمومی در وب وجود دارد که هرروز بر این تعداد افزوده می‌شود. به دلیل وجود چندین موتور کاوش برای بازیابی اطلاعات، موضوع ارزیابی کارآمدی موتورهای کاوش به یکی از مهم‌ترین حوزه‌های پژوهشی درزمینهٔ بازیابی اطلاعات شده است. پژوهشگرانی نظیر واگان[2] (۲۰۰۴) و لوان‌دافسکی[3] (۲۰۱۲) سنجه‌های مختلفی را برای ارزیابی موتورهای کاوش ارائه داده‌اند که هرکدام از آن‌ها ویژگی‌ها و مشخصه‌های خاصی از موتورهای کاوش را موردتوجه قرار می‌دهند. برای پژوهشگران حوزه ارزیابی موتورهای کاوش، آگاهی از سنجه‌های مختلف ارزیابی اهمیت فراوانی دارد؛ ازاین‌رو در پژوهش‌هایی نظیر یوسفی (۱۳۷۶)، پائو (۲۰۰۰)، پاورز[4] (۲۰۱۱)، حریری، باب‌الحوائجی، فرزندی پور و نادی راوندی (۱۳۹۳) و باما، احمد و سروانان[5] (۲۰۱۵) به‌مرور و تحلیل این سنجه‌ها پرداخته شده است. باوجوداین، سنجه‌های جدیدی توسط پژوهشگران برجسته حوزه علوم اطلاعات نظیر میزاروف[6] (۲۰۰۱)، یارولین و کامپولین[7] (۲۰۰۲) مطرح شده که در پژوهش‌های ارزیابی موتورهای کاوش از آن‌ها غفلت شده است و اطلاعات چندانی از این سنجه‌ها در دست نیست؛ درحالی‌که آگاهی از این سنجه‌ها و شناخت آن‌ها برای به‌کارگیری در ارزیابی، به‌عنوان پراستفاده‌ترین نظام‌های اطلاعاتی، اهمیت دارد. پژوهشی که در یکجا ضمن فهرست کردن آن‌ها به‌مرور سنجه‌های ارزیابی موتورهای کاوش و تحلیل ویژگی‌های آن‌ها بپردازد، صورت نگرفته است. در همین راستا، هدف از این پژوهش آن است تا سنجه‌های مختلف ارزیابی موتورهای کاوش فهرست، مرور و ویژگی‌های آن‌ها به‌منظور شناسایی در راستای استفاده از موتورهای کاوش، تحلیل شود.

روش‌شناسی

این پژوهش، مطالعه‌ای مروری است که در آن ۱۱ منبع فارسی و ۳۲ منبع انگلیسی، با دیدی انتقادی به بررسی و تحلیل پژوهش‌های سنجه‌های ارزیابی موتورهای کاوش پرداخته شده است. برای رسیدن به این هدف، ابتدا لازم بود تا مقاله‌های مرتبط در این زمینه شناسایی و مطالعه شود. بدین منظور، عبارت‌های جستجوی «سنجه‌های ارزیابی»، «ارزیابی موتورهای کاوش»، «ارزیابی موتورهای جستجو» «ارزیابی نظام بازیابی اطلاعات» و «سنجه‌های ارزیابی ربط» در پایگاه‌های اطلاعاتی مگ ایران[8] و سید[9] جستجو شد؛ همچنین عبارت‌های جستجوی « evaluation search engine»، « evaluation measure or metrics information retrival»، « method and information retrieval» و «evaluation measures or metrics» در موتور کاوش گوگل اسکالر[10] جستجو شد و سپس مقالات بازیابی شده مرتبط مطالعه شدند.

منابع مطالعه شده برای انجام این پژوهش در دو زبان فارسی و انگلیسی بودند که در جدول ۱ به‌صورت مختصر اشاره‌شده است.

جدول ۱. منابع موردمطالعه جهت انجام پژوهش

ردیف	سنجه ارزیابی	منابع
۱	دقت	آزادی (۱۳۸۴)، محمد اسماعیل و همکاران (۱۳۸۷)، بابایی و ساجدی (۱۳۹۲)، حریری و وکیلی مفرد (۱۳۹۲)، قاسمی و همکاران (۱۳۹۴)، ریاحی نیا و همکاران (۱۳۹۵)، محمد اسماعیل و نراقیان (۱۳۹۶)، عباسی دشتکی و چشمه سهرابی (۱۳۹۸)
		بارایلان[11] (۱۹۹۸)، شانگ و لی[12] (۲۰۰۲)، شافی و راثر[13] (۲۰۰۵)، دیمرسی[14] و همکاران (۲۰۰۷)، لوپز و ریبه رو[15] (۲۰۱۱)، حریری (۲۰۱۱)، گارفالو[16] (۲۰۱۲)، زویا[17] و زویا (۲۰۱۲)، زینالی تازه‌کندی و نوکاریزی (۲۰۲۰)
۲	بازیافت	بابایی و ساجدی (۱۳۹۲)، کوشا (۱۳۸۲)
		کنت[18] و همکاران (۱۹۵۵)، لنکستر[19] (۱۹۷۹)، کلارک و ویلت[20] (۱۹۹۷)، لاندونی و بل[21] (۲۰۰۰)، بیلال[22] (۲۰۱۲)، نوکاریزی و زینالی تازه کندی (۲۰۱۹)،
۳	اتحاد جاکارد	منبع فارسی مشاهده نشد
		بورلند و اینگورسن[23] (۱۹۹۸)، بورلند (۲۰۰۳)
۴	اتحاد کسینوسی	منبع فارسی مشاهده نشد
		بورلند و اینگورسن (۱۹۹۸)، بورلند (۲۰۰۳)

[1] Www.thesearchenginelist.com
[2] Vaughan
[3] Lewandowski
[4] Powers
[5] Bama, Ahmed and Saravanan
[6] Mizzaro
[7] Järvelin and Kekäläinen

[8] https://www.magiran.com/
[9] https://www.sid.ir/fa/journal/
[10] https://scholar.google.com/
[11] Bar-Ilan
[12] Shang & Li
[13] Shafi and Rather,
[14] Demirci
[15] Lopes & Ribeiro
[16] Garoufallou
[17] Zuya
[18] Kent
[19] Lancaster
[20] Clarke and Willett,
[21] Landoni, M. and Bell,
[22] Bilal
[23] Borlund & Ingwersen

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

متوسط فاصله	منبع فارسی مشاهده نشد میزارو[1] (۲۰۰۱)، میا و میزارو (۲۰۰۴)، میا و همکاران (۲۰۰۶)، زاهو و یائو[2] (۲۰۱۰)، سیرتکین (۲۰۱۲)
سود تجمعی تعدیل‌یافته نرمال	گل زردی و همکاران (۱۳۹۲)، یارولین و کامپولین (۲۰۰۲)، المسکری[3] و همکاران (۲۰۰۷)، ساکی و سونگ[4] (۲۰۱۱)، ساکی (۲۰۱۲)، سیرتکین (۲۰۱۲)،
بی پرف	منبع فارسی مشاهده نشد باکلی و وهیز[5] (۲۰۰۴)، سیرتکین (۲۰۱۲)
کارآمدی رتبه	منبع فارسی مشاهده نشد. گرنکوئیست[6] (۲۰۰۶)
جامعیت	محمد اسماعیل و همکاران (۱۳۸۷)، رجبی و نوروزی (۱۳۹۴)، ریاحی نیا و همکاران (۱۳۹۵)، محمد اسماعیل و نراقیان (۱۳۹۶)، فریک[7] (۱۹۹۸)، اسمیت[8] (۲۰۰۳)، شافی و راثر (۲۰۰۵)، کومار و پراکاش[9] (۲۰۰۹)، کومار و پاویترا[10] (۲۰۱۰)، عثمانی[11] و همکاران (۲۰۱۲)، کومار و باهادو (۲۰۱۳)، نوکاریزی و زینالی تازه کندی (۲۰۱۹)، زینالی تازه کندی و نوکاریزی (۲۰۲۰)

همان‌گونه که در جدول ۱ مشاهده می‌شود، می‌توان گفت که سنجه دقت پرکاربردترین سنجه برای ارزیابی موتورهای کاوش بوده است. افزون بر این، سنجه‌های جدید ارزیابی ناشناخته مانده‌اند؛ به‌ویژه در پژوهش‌های فارسی اغلب این سنجه‌ها مورد غفلت واقع‌شده است. ازاین‌رو، در این پژوهش سنجه‌های ارزیابی موتورهای کاوش بدون محدودیت زمانی، تحلیل و درنهایت جوانب مختلف موضوع مشخص و درنهایت رابطه بین سنجه‌ها و معماری موتورهای کاوش تبیین شده است.

سنجه‌های ارزیابی

سنجه‌های مختلفی به‌منظور ارزیابی موتورهای کاوش ارائه شده است. آنچه در این رابطه حائز اهمیت است، توجه به تمایز بین دو مفهوم کارآمدی و کارایی است. بر اساس نظر گول و یاداو[12] (۲۰۱۲) و کرافت و همکاران (۲۰۱۵) کارآمدی، توانایی یک موتور کاوش را در بازیابی مدارک مرتبط می‌سنجد؛ درحالی‌که کارایی سرعت یک موتور کاوش را در انجام بازیابی می‌سنجد. برای اینکه یک موتور کاوش به‌عنوان موتور کارآمد و کارا شناخته شود؛ بایستی همه اجزا آن به نحو نیکو عمل نمایند. به‌بیان‌دیگر نحوه عملکرد هرکدام از اجزا بر نتیجه نهایی یعنی همان مدارک بازیابی شده تأثیر دارد. به‌هرحال در ادامه به گروه‌بندی سنجه‌های ارزیابی کارآمدی –نه کارایی– موتورهای کاوش پرداخته شده است.

سنجه‌های غیر رتبه‌بندی

سنجه‌های مختلفی برای ارزیابی کارآمدی موتورهای کاوش وجود دارد که هرکدام از آن‌ها ویژگی خاصی از آن را موردسنجش قرار می‌دهند. در حالت کلی، می‌توان گفت که بیشتر سنجه‌های که هم‌زمان با مطرح‌شدن مباحث ارزیابی پیشنهادشده است سنجه‌هایی نظیر دقت، بازیافت و نظیر آن هستند که به نحوه‌ی رتبه‌بندی مدارک توجه نشده است. به‌بیان‌دیگر، سنجه‌هایی در این گروه قرار می‌گیرند که ترتیب بازیابی مدارک توسط موتورهای کاوش بر نمره به‌دست‌آمده از سنجه تأثیر ندارد که در ادامه بدان‌ها اشاره‌شده است.

دقت

احتمالاً سنجه دقت قدیمی‌ترین سنجه ارزیابی است که در سال ۱۹۶۶ به‌عنوان بخشی از پروژه کرانفیلد[13] طراحی شد (کلوردن و کین[14]، ۱۹۶۸ نقل در سیرتکین[15]، ۲۰۱۲). بر اساس نظر ساراسویک[16]، (۲۰۱۵) در اواسط دهه ۱۹۵۰ آلن کنت[17] (۱۹۲۲–۲۰۱۴) و جیمز دابلیو پری[18] (۱۹۰۷–۱۹۷۱) از افراد مطرح و پیشگام در علم اطلاعات، مجموعه مقاله‌هایی را درباره فنون بازیابی اطلاعات نوشتند که در یکی از این مقاله‌ها، سنجه‌هایی را برای ارزیابی پیشنهاد دادند که یکی از این سنجه‌ها، سنجه دقت بود. دقت که با عنوان اعتماد[19] نیز شناخته می‌شود به کسری از مدارک بازیابی شده مرتبط اشاره دارد. همچنین می‌توان گفت که دقت به کسری از نمونه‌های پیش‌بینی‌شده مثبت اشاره دارد که واقعاً مثبت هستند (حریری و همکاران، ۱۳۹۳؛ پاورز، ۲۰۱۱) که از طریق فرمول زیر قابل‌محاسبه است $P = \frac{|A \cap B|}{|B|}$ در این فرمول منظور از p، میزان دقت؛ منظور از A، مجموعه مدارک مرتبط در پایگاه؛ منظور از B، مجموعه مدارک بازیابی شده است. علامت ∩، به اشتراک دو مجموعه یادشده اشاره دارد. در این فرمول، A∩B به مجموعه مدارکی اشاره دارد که هم مرتبط و هم بازیابی

[1] Mizzaro
[2] Zhou& Yao
[3] Al-Maskar
[4] Sakai & Song,
[5] Buckley & Voorhees
[6] Grönqvist
[7] Frické
[8] smith
[9] Kumar and Prakash,
[10] Kumar and Pavithra,
[11] Usmani
[12] Goel & Yadav
[13] Cranfield Project
[14] Cleverdon and Keen
[15] Sirotkin
[16] Saracevic
[17] Allen Kent
[18] James W. Perry
[19] Confidence

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

شده باشند. به‌بیان‌دیگر دقت، نسبت تعداد مدارک بازیابی شده مرتبط به تعداد کل مدارک بازیابی شده است. برای مثال اگر در یک جستجو برای یک پرسش، ۸ مدرک بازیابی شده است که از بین آن‌ها، ۴ مدرک مرتبط باشد، در این صورت خواهیم داشت:

$$\text{دقت} = \frac{۴}{۸}$$

در رابطه با این سنجه می‌توان گفت که دقت به مدارک بازیابی شده مرتبط اشاره دارد (سیرتکین، ۲۰۱۲). به‌بیان‌دیگر، سنجه دقت به توانایی موتور کاوش در رابطه با بازیابی نکردن مدارک نامرتبط اشاره دارد (کرافت، متزلر و استروهمن، ۲۰۱۵) و مدل هنجاری سنجه دقت این است که مدارک بازیابی شده، مرتبط باشند؛ لذا مدارک نامرتبط بازیابی نشوند. ازاین‌رو، اگر دو موتور کاوش الف و ب به ترتیب ۸ و ۶ مدرک را بازیابی نمایند که از بین آن‌ها، به ترتیب ۴ و ۳ مدرک مرتبط باشند، در این صورت دقت هر دو موتور کاوش الف و ب به یک‌میزان و برابر با ۰.۵ خواهد بود؛ درحالی‌که موتور کاوش الف، ۴ مدرک مرتبط و موتور کاوش ب، ۳ مدرک مرتبط بازیابی کرده است. بااین‌وجود، دقت هر دو موتور کاوش یکسان است و این ضعف سنجه دقت است. افزون بر این، در این سنجه به مدارک مرتبط موجود در پایگاه نمایه موتور کاوش و وب توجه نشده است.

بازیافت

سنجه بازیافت، یکی دیگر از سنجه‌هایی است که در بیشتر پژوهش‌های ارزیابی موتورهای کاوش مورداستفاده قرارگرفته است. بازیافت عبارت است از میزان مدارک مربوطی که عملاً از میان کل مجموعه مدارک مربوط در فایل بازیابی شده است (پائو، ۲۰۰۰). این سنجه نیز توسط کنت و پری ارائه‌شده است (ساراسویک، ۲۰۱۵) که توانایی موتور کاوش در بازیابی مدارک مرتبط موجود در پایگاه را نشان می‌دهد (کلارک و ویلت[1]، ۱۹۹۷؛ ییلماز، کارتویت و کان اویلس[2]، ۲۰۱۲) که از فرمول زیر قابل‌محاسبه است:

$$R = \frac{|A \cap B|}{|A|}$$

در این فرمول، منظور از R، میزان بازیافت است و A∩B به مجموعه مدارکی اشاره دارد که هم مرتبط و هم بازیابی شده باشند. ازلحاظ مفهومی بازیافت به کسری از مدارک مرتبط اشاره دارد که بازیابی شده است (کرافت و همکاران، ۲۰۱۵). به‌طور فرضی اگر در پایگاه اطلاعاتی یک موتور کاوش در رابطه با یک پرسش ۲۰ مدرک مرتبط وجود دارد که در جستجوی انجام‌شده فقط ۵ مدرک مرتبط بازیابی شده است، در این صورت خواهیم داشت:

در محاسبه سنجه دقت، در مخرج کسر اندازه مجموعه مدارک بازیابی شده قرار می‌گیرد؛ درحالی‌که در محاسبه سنجه بازیافت، در مخرج کسر اندازه مجموعه مدارک مرتبط موجود در پایگاه قرار می‌گیرد. اگر تعداد مدارک مرتبط موجود در پایگاه دو موتور کاوش الف و ب برابر ۱۶ باشد، در این صورت میزان بازیافت موتور کاوش الف و ب به ترتیب ۰/۲۵ و ۰/۱۸ خواهد بود. با مقایسه دو سنجه دقت و بازیافت مشخص می‌شود که سنجه دقت به تعداد کل مدارک موجود در پایگاه اطلاعاتی موتور کاوش حساس نیست؛ درحالی‌که سنجه بازیافت، تعداد کل مدارک اندازه پایگاه اطلاعاتی موتور کاوش را نیز در نظر می‌گیرد. بااین‌وجود، باید اشاره شود که مدل هنجاری سنجه بازیافت این است که مدارک مرتبطی که در پایگاه موتور کاوش نمایه شده است، به هنگام جستجوی کاربران بازیابی شود؛ درحالی‌که کاربران مایل‌اند به کلیه مدارک مرتبط موجود در وب دسترسی پیدا کنند (نوکاریزی و زینالی تازه‌کندی، ۲۰۱۹).

در رابطه با سنجه بازیافت، ساراسویک (۲۰۱۵) بیان می‌دارد، این سنجه ابتدا توسط کنت و پری مطرح شد که ربط[3] نام داشت و سپس عنوان بازیافت به خود گرفته است. در همین راستا، پائو (۲۰۰۰) نیز از سنجه‌ای به نام ربط نام‌برده است. افزون بر این، ساراسویک (۲۰۱۵) معتقد است که این سنجه مبهم است. در همین راستا، نوکاریزی و زینالی تازه‌کندی (۲۰۱۹) تأکید می‌کنند که سنجه بازیافت نیاز به بازاندیشی دارد و سنجه جدیدی را با عنوان جامعیت مطرح و بین بازیافت و جامعیت تمایز قائل شدند. بهتر است اشاره شود که برخی از پژوهشگران نظیر پاورز (۲۰۱۱) از عنوان حساسیت[4] نیز برای اشاره به سنجه بازیافت استفاده کرده‌اند.

جامعیت[5]

بر اساس نظر نوکاریزی و زینالی تازه‌کندی (۲۰۱۹)، سنجه بازیافت به نحوه سازمان‌دهی مدارک اشاره دارد؛ درحالی‌که کاربران تمایلی به مسائل فنی موتورهای کاوش ندارند؛ بلکه آن‌ها مایل‌اند به مدارک مرتبط موجود در وب دست یابند. برای روشن شدن مطلب بهتر است اشاره شود که واژه بازیافت از دو واژه باز و یافتن تشکیل‌شده است. توجه به ساختمان ادبی خود روشنگر مطلب است که یک‌بار مدارک از وب یافت شده و در موتور کاوش نمایه‌سازی شده‌اند (مرحله یافت[6]) و وقتی کاربری در موتور کاوش به جستجو می‌پردازد، این مدارک یافت شده از وب یعنی نمایه شده در پایگاه اطلاعاتی موتور کاوش توسط الگوریتم بازیابی موتور کاوش

[1] Clarke & Willett
[2] Yilmaz, Carterette & Kanoulas
[3] relevance
[4] Sensitivity
[5] comprehensiveness
[6] call

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

بررسی‌شده و مدارکی از بین آن‌ها برای کاربر ارائه می‌شود (مرحله بازیافت). بر همین اساس سنجه بازیافت به وب توجه ندارد؛ بلکه به پایگاه اطلاعاتی موتور کاوش–مدارک موجود در پایگاه موتور کاوش– تمرکز دارد. درصورتی‌که کاربران به دنبال مدارک مرتبط در وب هستند نه مدارک مرتبطی که در پایگاه موتور کاوش نمایه شده‌اند. در همین راستا، نوکاریزی و زینالی تازه‌کندی (۲۰۱۹) سنجه جامعیت را مطرح کرده‌اند که به کسری از مدارک مرتبط وب اشاره دارد که توسط موتورهای کاوش بازیابی شده است که فرمول آن به‌صورت زیر تعریف شده است:

$$C_1 = \frac{|A_1 \cap B|}{|(A_1 \cap B) \ U \ (A_2 \ \cap B) U \ldots, , U \ (A_n \cap B)|}$$

در این فرمول، منظور از C، میزان جامعیت؛ منظور A_1، مجموعه مدارک مرتبط موجود در پایگاه موتور کاوش یک؛ منظور از A_2، مجموعه مدارک مرتبط موجود در پایگاه موتور کاوش دو؛ و منظور از A_n، مجموعه مدارک مرتبط موجود در پایگاه موتور کاوش n؛ و منظور از B، مجموعه مدارک بازیابی شده است. ازاین‌رو منظور از $A_1 \cap B$، مجموعه مدارک مرتبط موجود در موتور کاوش یک است که بازیابی شده است.

برای مقایسه سنجه بازیافت و جامعیت استفاده از مثال قبلاً مطرح‌شده روشنگر مطلب خواهد بود. اگر در رابطه با یک پرسش در وب، ۴۰ مدرک مرتبط وجود داشته باشد که از این تعداد، ۲۰ مدرک توسط یک موتور کاوش نمایه‌سازی شده است. کاربری در این موتور کاوش به جستجو پرداخته و فقط ۵ مدرک مرتبط بازیابی شده است، در این صورت در محاسبه سنجه جامعیت خواهیم داشت:

در مورد همین مثال مطرح‌شده، نمره سنجه بازیافت ۰.۲۵ بود؛ درحالی‌که نمره سنجه جامعیت ۰.۱۲ به‌دست‌آمده است. به همین سبب استفاده از هرکدام از این دو سنجه، نمره متفاوتی برای کارآمدی موتورهای کاوش به دست خواهد داد. افزون بر این چه‌بسا امکان دارد که یک موتور کاوش در مقایسه با موتور کاوشی دیگر نمره بازیافت بیشتری کسب نماید؛ اما در حقیقت کارآمدی این موتور کاوش از کارآمدی موتور کاوش دیگر کم‌تر باشد. به‌بیان‌دیگر، سنجه جامعیت، سنجه قوی‌تری از سنجه بازیافت است که برای نشان دادن این امر مثالی در جدول ۱ ارائه شده است.

جدول ۱. مقایسه دو سنجه بازیافت و جامعیت

تعداد کل مدارک مرتبط موجود در وب	۴۰	بازیافت موتور الف	۰.۶
تعداد مدارک مرتبط موجود در موتور الف	۵	جامعیت موتور الف	۰.۰۷
تعداد مدارک مرتبط بازیابی شده از موتور الف	۳		
تعداد مدارک مرتبط موجود در موتور ب	۲۰	بازیافت موتور ب	۰.۵
تعداد مدارک مرتبط بازیابی شده از موتور ب	۱۰	جامعیت موتور ب	۰.۲۵

همان‌گونه که در جدول ۱ مشاهده می‌شود، ۵ مدرک مرتبط توسط موتور الف و ۱۰ مدرک مرتبط توسط موتور ب بازیابی شده است. باوجوداین، نمره بازیافت موتور الف (۳/۵) بیشتر از نمره بازیافت موتور ب (۱۰/۲۰) است که کارآمدی دو موتور را به‌صورت درست نشان نمی‌دهد؛ درحالی‌که نمره جامعیت موتور ب (۱۰/۴۰) تفاوت فراوانی با نمره جامعیت موتور الف (۳/۴۰) دارد که کارآمدی دو موتور را به‌خوبی نشان می‌دهد. بدین ترتیب می‌توان گفت که سنجه جامعیت در مقایسه با سنجه بازیافت به‌صورت دقیق‌تری نشان‌دهنده کارآمدی موتورهای جستجو است.

سنجه‌های رتبه‌بندی

ضعف اساسی همه سنجه‌های غیر رتبه‌بندی این است که به جایگاه بازیابی مدارک حساس نیست. به‌بیان‌دیگر، چه مدرک اول مرتبط و مدرک دوم نامرتبط باشد و یا مدرک اول نامرتبط و مدرک دوم مرتبط باشد؛ در این سنجه‌ها، کارآمدی موتورهای کاوش یکی پنداشته می‌شود. ازاین‌رو، پژوهشگران مختلف حوزه ارزیابی موتورهای کاوش با توجه به این امر شروع به طراحی سنجه‌هایی نمودند که به ترتیب مدارک نیز حساس باشد؛ زیرا در حقیقت موتور کاوشی کارآمد هست که مدارک مرتبط را قبل از مدارک نامرتبط بازیابی نماید. افزون بر این موتوری که مدارک مرتبط‌تر را قبل از مدارک مرتبط بازیابی نماید، از کارآمدی بیشتری برخوردار است. در همین راستا، سنجه‌های رتبه‌بندی مطرح شدند که در ادامه به ترتیب با توجه به زمان مطرح‌شدن این سنجه‌ها به آن‌ها اشاره‌شده است.

اتحاد جاکارد[1]

بر اساس نظر بورلند و اینگورسن (۱۹۹۸)، در ارزیابی موتورهای کاوش تعاملی، چالشی با عنوان ماهیت پویای نیاز اطلاعاتی تأییدشده است. ازاین‌رو آگاهی از چندبعدی و چندوجهی بودن ربط، نحوه ارزیابی موتورهای کاوش را نیز تغییر داده است که این عامل خود سبب تغییر نحوه سنجش و آزمون آن‌ها شده است. ازآنجایی‌که در متونی نظیر داورپناه (۱۳۸۳)، اینگورسن (۲۰۱۰) و ساراسویک (۲۰۱۵) از دو رویکرد نظام‌گرا و کاربرگرا یادشده است.

[1] Jaccard Association

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

ازاین‌رو، بورلند و اینگورسن[1] (۱۹۹۸) ادعا می‌کنند که یک راه‌حل عملی برای ایجاد پل برای شکاف موجود بین ربط عینی و ذهنی ارائه کرده‌اند. آن‌ها در این پژوهش، دو سنجه برای ارزیابی کارآمدی موتورهای کاوش پیشنهاد دادند که یکی از این سنجه‌ها، سنجه اتحاد جاکارد است. بورلند و اینگورسن (۱۹۹۸) معتقدند که ربط، مفهومی چندوجهی و چندبعدی است؛ لذا سنجه آن‌ها نیز سنجه چندبعدی است که از طریق فرمول زیر قابل‌محاسبه است:

$$Jaccard\ Association = \frac{|R_1 \cap R_2|}{|R_1 U R_2|}$$

سنجه اتحاد جاکارد رابطه بین دو مجموعه از موجودیت‌ها را نشان می‌دهد که در ارزیابی نظام‌های بازیابی اطلاعات به رابطه برونداد دو نوع از ارزیابی ربط اشاره دارد. در این فرمول، منظور از R_1،R_2 ارزش ربط اختصاص داده‌شده به مدارک توسط متخصص، کاربر یا نظام بازیابی اطلاعات است.

همان‌گونه که در جدول ۲ مشاهده می‌شود، اتحاد دو نوع ربط از طریق دو سنجه اتحاد جاکارد و اتحاد کسینوسی محاسبه‌شده است که تفاوت زیادی در نمره این دو سنجه وجود دارد. این تفاوت ناشی از مخرج کسر این دو سنجه است. با مقایسه دو سنجه اتحاد جاکارد و اتحاد کسینوسی به نظر می‌رسد که سنجه اتحاد کسینوسی سنجه قوی‌تری از اتحاد جاکارد باشد.

جدول ۲. محاسبه سنجه شاخص جاکارد و سنجه اتحاد کسینوسی

مدارک	الف	ب	پ	ت	ج
نمره ربط نوع اول	۰.۹	۰.۸	۰.۸	۰.۷	۰.۵
نمره ربط نوع دوم	۰.۹	۰.۸	۰.۸	۰.۷	۰.۵
شاخص جاکارد	۰.۶۱۹				
اتحاد کسینوسی	۱				

اتحاد کسینوسی[3]

اتحاد کسینوسی، یکی دیگر از سنجه‌هایی است که توسط بورلند و اینگورسن (۱۹۹۸) برای توجه به دو رویکرد موجود در ارزیابی موتورهای کاوش ارائه شده است. ازنظر آن‌ها این سنجه نیز، سنجه چندبعدی است که با استفاده از آن اتحاد بین ربط عینی و ربط ذهنی برقرار می‌شود. سنجه اتحاد کسینوسی به طریق زیر قابل‌محاسبه است:

$$Cosine\ Association = \frac{\Sigma R_1 R_2}{\sqrt[2]{(\Sigma {R_1}^2) * (\Sigma {R_2}^2)}}$$

در فرمول ذکرشده منظور از R_1، ارزش ربط اختصاص داده‌شده به مدارک توسط موتور کاوش و منظور از R_2، ارزش ربط اختصاص داده‌شده به مدارک توسط کاربر یا متخصص هست. برای مقایسه دو سنجه اتحاد جاکارد و اتحاد کسینوسی مثالی در جدول ۲ ارائه شده است.

متوسط فاصله[2]

بر اساس نظر میزارو (۲۰۰۱) سنجه‌های دقت و بازیافت دارای نقص‌هایی بسیاری هستند؛ ازجمله اینکه در محاسبه سنجه دقت و بازیافت قضاوت ربط دودویی یا صفر و یک است؛ درحالی‌که در حقیقت، ربط مدارک به‌صورت مقیاسی پیوسته است که می‌تواند اعداد اعشاری را نیز بپذیرد. نقص دوم دو سنجه یادشده این هست که دارای مدل هنجاری هستند؛ به‌بیان‌دیگر، از قبل معیارهای را برای مطلوبیت موتورهای کاوش در نظر گرفته‌اند؛ درحالی‌که مطلوبیت آن‌ها فقط باید توسط کاربران تعیین شود. سومین نقص به حساس نبودن این دو سنجه به مدارک نامرتبط بازیابی نشده اشاره دارد؛ همچنین در این دو سنجه به رتبه‌بندی مدارک یعنی ترتیب نمایش مدارک توجه نشده است. ازاین‌رو، میزارو سنجه جدیدی با عنوان متوسط فاصله فرمول‌بندی کرده است. این سنجه، فاصله بین ربط معیار مدارک و ربط پیش‌بینی‌شده یا اختصاص داده‌شده به مدارک توسط موتور کاوش را نشان می‌دهد که با استفاده از فرمول زیر قابل‌محاسبه است:

$$ADM = 1 - \frac{\Sigma |SRE - URE|}{|D|}$$

در این فرمول منظور از ADM، سنجه متوسط فاصله؛ SRE، نمره ربط اختصاص داده‌شده به مدارک توسط موتور کاوش و URE، نمره ربط تعیین‌شده مدارک از سوی کاربران؛ D، تعداد مدارک موجود در پایگاه است و علامت $\boldsymbol{\Sigma}$ نشان‌دهنده تجمیع است. سنجه متوسط فاصله، نمره‌ای بین صفر و یک است که نمره صفر نشان‌دهنده عملکرد ضعیف موتور کاوش و نمره یک، بهترین عملکرد آن را نشان می‌دهد (میا[4] و میزارو، ۲۰۰۴). برای آشنایی بیشتر در رابطه با تفاوت سنجه متوسط فاصله، دقت و بازیافت، پنج مدرک فرضی همراه با نمره ربط اختصاص داده‌شده توسط سه

[1] Borlund & Ingwersen
[2] Average Distance
[3] Cosine Association
[4] Mea

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

موتور کاوش فرضی و نمره ربط تعیین‌شده از سوی کاربران و میزان سنجه‌های یادشده در جدول ۳ ارائه شده است.

جدول ۳. محاسبه سنجه متوسط فاصله

مدارک	۱	۲	۳	۴	۵	دقت	بازیافت	متوسط فاصله
نمره ربط کاربران	۰.۸	۰.۶	۰.۴	۰.۲	۰.۱			
نمره موتور ۱	۰.۹	۰.۵	۰.۵	۰.۱	۰.۲	۰.۶۷	۰.۸۴	۰.۹
نمره موتور ۲	۱	۰.۴	۰.۶	۰	۰.۳	۰.۵	۰.۵	۰.۸
نمره موتور ۳	۰.۸	۰.۶	۰.۴	۰.۲	۱	۰.۶۷	۰.۸۴	۰.۸

در جدول ۳، برای مقایسه سه سنجه دقت، بازیافت و متوسط فاصله از پیش‌فرض ارزش بیشتر مساوی ۰.۵ برای تصمیم‌گیری بین مرتبط و غیر مرتبط و بازیابی شده و بازیابی نشده استفاده‌شده است. حال برای مقایسه سه سنجه به جدول ۳ توجه نمایید. با توجه به جدول ۳ مشخص است که موتور اول بهتر از موتور دوم عمل کرده است؛ زیرا نمره ربط اختصاص داده‌شده به مدارک توسط موتور اول در مقایسه با نمره ربط اختصاص داده‌شده به مدارک توسط موتور دوم، نزدیک‌تر و مشابه‌تر به نمره ربط اختصاص داده‌شده به مدارک توسط کاربران است، اما مقایسه موتور اول و موتور سوم مشکل است. موتور سوم در مقایسه با موتور اول در همه موارد –به‌جز مدرک پنجم– عملکرد بهتری دارد، اما نمره ربط اختصاص داده‌شده به مدرک پنجم توسط موتور سوم واقعاً نادرست است. میزان دقت و بازیافت نظام اول و سوم متفاوت نیست؛ درحالی‌که میزان متوسط فاصله دو موتور اول و سوم تفاوت دارند. به همین سبب می‌توان گفت که سنجه متوسط فاصله در مقایسه با دو سنجه دقت و بازیافت تفاوت موتورهای کاوش را به‌صورت بهتری نشان می‌دهد. ازنظر میزارو (۲۰۰۱) سنجه متوسط فاصله، ویژگی‌های یک سنجه خوب تعیین‌شده (توسط سویت[1]، ۱۹۷۳ نقل در میزارو ۲۰۰۱) را دارا است. ازنظر وی سنجه یادشده، فقط کارآمدی موتور کاوش را می‌سنجد، آن قدرت تمایز گذاری آن را مشخص می‌کند، فاقد معیارهای از پیش تعیین‌شده است و هیچ پیش فرضی خاصی ندارد و آن به‌طور کامل از قضاوت کاربران پیروی می‌کند.

سود تجمعی تعدیل‌یافته نرمال[2]

محیط‌های بازیابی مدرن بزرگ، با استفاده از خروجی بزرگ خود، کاربران خود را دست‌پاچه می‌کنند. ازآنجایی‌که همه مدارک به‌اندازه یکسان به کاربران مفید نیستند؛ بنابراین باید ابتدا مدارک با ربط بالا به کاربران نمایش داده شود. افزون بر این، در بسیاری از مواقع، کاربران تمایل دارند، در جستجوی خود فقط نتایج اولیه بازیابی شده توسط موتورهای کاوش را بررسی نمایند. در چنین مواقعی، سنجه‌های بازیافت و دقت سنجه‌های مناسبی نیستند. در همین راستا، یارولین و کامپولین (۲۰۰۲) معتقدند که تمرکز سنجه‌های کارآمدی باید بر این باشد که موتورهای کاوش تا چه اندازه می‌توانند مدارک مرتبط‌تر را جلوتر از مدارک مرتبط بازیابی نمایند. ازاین‌رو آن‌ها سنجه جدیدی را با عنوان سود تجمعی تعدیل‌یافته نرمال مطرح کرده‌اند که فرمول آن به شکل زیر است:

$$NDCG = \frac{DCG}{IDCG}$$

در این فرمول NDCG مخفف Normalized Discounted Cumulative Gain بوده سود تجمعی تعدیل‌یافته نرمال، DCG مخفف Discounted Cumulative Gain بوده و سود تجمعی تعدیل‌یافته و IDCG مخفف Ideal Discounted Cumulative Gain و سود تجمعی تعدیل‌یافته ایده آل است. برای روشن شدن بیشتر در رابطه با نحوه محاسبه این سنجه، ارائه مثالی (جدول ۴) مفید خواهد بود. فرض کنید که به ترتیب (از راست به چپ) چهار مدرک زیر توسط یک موتور کاوش بازیابی شده است:

$\frac{۲}{۵}$ $\frac{۱}{۵}$ $\frac{۳}{۵}$ $\frac{۱}{۵}$

جدول ۴. مراحل محاسبه سود تجمعی تعدیل یافته نرمال

مراحل	مفاهیم	محاسبات
مرحله اول	محاسبه لگاریتم رتبه مدارک در مبنای ۲	۱ و ۱ و ۱/۵۸ و ۲
مرحله دوم	تقسیم میزان ربط مدارک بر اعداد حاصل از مرحله اول	$\frac{۱}{۵}$ $\frac{۳}{۵}$ $\frac{۱}{۵}$ $\frac{۲}{۵}$ / ۱ ۱ ۵۸/۱ ۲
مرحله سوم	مرتب‌سازی مدارک بر اساس میزان ربطشان (حالت مطلوب)	$\frac{۳}{۵}$ $\frac{۲}{۵}$ $\frac{۱}{۵}$ $\frac{۱}{۵}$
مرحله چهارم	محاسبه لگاریتم رتبه مدارک در مبنای ۲ برای مرحله سوم	۱ و ۱ و ۱/۵۸ و ۲
مرحله پنجم	تقسیم اعداد حاصل از مرحله سوم بر اعداد حاصل از مرحله چهارم	$\frac{۳}{۵}$ $\frac{۲}{۵}$ $\frac{۱}{۵}$ $\frac{۱}{۵}$ / ۱ ۱ ۵۸/۱ ۲

درنهایت، اعداد حاصل از مرحله دوم بر اعداد حاصل از مرحله پنجم تقسیم خواهد شد که نتیجه حاصل، سود تجمعی تعدیل‌یافته نرمال را نشان می‌دهد.

[1] Swets
[2] normalized discounted cumulative gain

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

بی پرف[1]

بر اساس نظر باکلی و وهیز[2] (۲۰۰۴) سنجه‌های دقت و بازیافت به‌عنوان سنجه‌های استاندارد در پژوهش‌های ارزیابی موتورهای کاوش استفاده می‌شود؛ اما باوجوداین، سنجه‌های یادشده فقط مبتنی بر مدارک قضاوت شده هستند و مدارک قضاوت نشده نادیده گرفته‌شده است. سنجه متوسط فاصله نیز از این امر مستثنی نیست. به‌بیان‌دیگر این سنجه‌ها بین مدارکی که نامرتبط تعیین‌شده‌اند و مدارکی که نامرتبط فرض شده‌اند، به دلیل اینکه این مدارک قضاوت نشده‌اند، تمایز قائل نمی‌شود. آن‌ها انگیزه خود را طراحی سنجه‌ای می‌دانند که مبتنی بر رابطه اولویت‌بندی و ترجیحی بوده و در مواجهه با قضاوت ناکامل مدارک قوی باشد. آن‌ها سنجه خود را بی پرف نام‌گذاری کردند زیرا سنجه یادشده برای تعیین رابطه ترجیحی مبتنی بر قضاوت ربط دودویی (صفر و یک) است. منظور از رابطه ترجیحی این است که کاربر یک مدرک را بر سایر مدارک در رابطه با یک موضوع یا نیاز اطلاعاتی ترجیح می‌دهد. سنجه یادشده از طریق فرمول زیر قابل‌محاسبه است:

$$bpref = \frac{1}{R} \Sigma \left(1 - \frac{|n \ ranked \ higher \ than \ r|}{R}\right)$$

در این فرمول منظور از bpref، سنجه بی پرف؛ R، تعداد مدارک مرتبط؛ n، تعداد مدارک نامرتبط و r، یک مدرک مرتبط؛ n ranked higher than r به تعداد مدارک نامرتبطی اشاره دارد که قبل از مدرک مرتبط موردنظر بازیابی شده است. در حالت ساده می‌توان بی پرف را به‌صورت حاصل‌ضرب تعداد مدارک مرتبط در تعداد مدارک نامرتبط تعریف نمود. برای آشنایی بیشتر با نحوه به‌کارگیری سنجه یادشده مثالی ارائه شده است. فرض کنید یک موتور کاوش فهرستی از مدارک را به ترتیب از راست به چپ به‌صورت زیر بازیابی نموده است:

لیست مدارک: ۱، ۰، ۱،۰، ۱

در این صورت سنجه بی پرف به‌صورت زیر به دست خواهد آمد:

$$b\,pref = \frac{1}{3}\left(\left(1-\frac{1}{3}\right) + \left(1-\frac{1}{3}\right) + \left(1-\frac{1}{3}\right)\right) = \frac{1}{3}\left(\frac{6}{3}\right) = \frac{6}{9} = ۰.۶۶$$

سنجه بی پرف عددی بین صفر و یک است که عدد یک، نهایت مطلوبیت موتور کاوش را در بازیابی مدارک مرتبط نشان می‌دهد. درنهایت باید اشاره شود، زمانی که تعداد مدارک مرتبط اندک باشد، استفاده از سنجه بی پرف اطلاعات مفیدی را برای مقایسه موتورهای کاوش ارائه نمی‌کند. بر اساس نظر سیرتکین (۲۰۱۲) در مقایسه سه سنجه متوسط فاصله، سود تجمعی تعدیل‌یافته نرمال و بی پرف بیان می‌دارد که دو سنجه متوسط فاصله و بی پرف مبتنی بر موتورهای کاوش هستند؛ درحالی‌که سنجه سود تجمعی تعدیل‌یافته نرمال به کاربران اهمیت بیشتری می‌دهد. افزون بر این به نظر ساکی[3] (۲۰۰۷) زمانی که مدارک قضاوت نشده نادیده گرفته می‌شود، سنجه سود تجمعی تعدیل‌یافته نرمال در مقایسه با سنجه بی پرف عملکرد موتورهای کاوش را به نحوه بهتر نشان می‌دهد. در پژوهشی دیگر، ساکی و کاندو[4] (۲۰۰۸) دو سنجه سود تجمعی تعدیل‌یافته نرمال و بی پرف را با استفاده از قضاوت ربط ناکامل مدارک مقایسه نموده‌اند. نتایج پژوهش آن‌ها نشان داد که سود تجمعی تعدیل‌یافته نرمال سنجه قوی‌تری از سنجه بی پرف است.

کارآمدی رتبه‌بندی[5]

در محاسبه سنجه بی‌پرف فقط تعداد مدارک مرتبط و تعداد مدارک نامرتبط قضاوت شده مورد استفاده قرار می‌گیرد و ازنظر گرنکویست[6] (۲۰۰۵)، سنجه بی پرف یک مشکل اساسی دارد و این مشکل به‌خصوص زمانی که تعداد مدارک مرتبط کم باشد، حادتر می‌شود؛ ازاین‌رو سنجه جدیدی به‌نام کارآمدی رتبه که مدارک قضاوت نشده را نیز در برمی‌گیرد، مطرح شد. به‌بیان‌دیگر، این سنجه از تعداد مدارک تأثیر می‌پذیرد. بوچرو، کلارک و کوماک[7] (۲۰۱۶) نیز خاطرنشان می‌سازند که برای مجموعه‌های پویا که به‌صورت دائم تعداد مدارک آن‌ها تغییر می‌کند (نظیر موتورهای کاوش)، استفاده از سنجه بی پرف در زمان‌های مختلف نتایج متفاوتی به دست خواهد داد و میزان پایداری نتایج حاصل از این سنجه کم‌تر است؛ به این دلیل که سنجه بی پرف به مدارک قضاوت نشده حساس نیست؛ درحالی‌که سنجه کارآمدی رتبه‌بندی، مدارک قضاوت نشده را نیز در نظر می‌گیرد. به همین سبب گرنکویست سنجه جدیدی با عنوان کارآمدی رتبه‌بندی طراحی نموده است که به طریق زیر قابل‌محاسبه است:

$$Rank\ eff = \frac{\Sigma I_{rr}(r)}{R(N-R)}$$

در این فرمول منظور از Rank eff، کارآمدی رتبه‌بندی؛ $I_{rr}(r)$، تعداد مدارک نامرتبط بازیابی شده قبل از مدرک مرتبط r؛ منظور از N، تعداد مدارک قضاوت شده؛ و منظور از R، تعداد مدارک مرتبط شناخته شده است. برای آشنایی با چگونگی محاسبه سنجه کارآمدی رتبه‌بندی، مثالی در جدول ۵ ارائه شده است. در این جدول منظور از علامت سؤال، مدارک قضاوت نشده است. فرض کنید که یکی از مدارک قضاوت نشده مرتبط است.

جدول ۵. محاسبه سنجه کارآمدی رتبه‌بندی

ربط مدارک	۱	؟	۰	۱	۰	؟	؟	؟	۱
سنجه کارآمدی رتبه‌بندی	۰.۶۲								

[1] b pref
[2] Buckley & Voorhees
[3] Sakai & kando
[4] Kando
[5] rank efficiency
[6] Grönqvist
[7] Büttcher, Clarke & Cormack

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

نتیجه‌گیری

فنون و سنجه‌های استفاده شده پژوهش‌های مرتبط با ارزیابی موتورهای کاوش، فعالیت‌هایی صرفاً فنی و بی‌ارتباط با مبانی نظری نیستند. این گفته بابیانی دیگر در پژوهش بورلند و اینگورسن (۱۹۹۸) ذکر گردیده است؛ زمانی که وی لزوم تغییر سنجه‌های ارزیابی را با توجه به تغییر نگرش در رابطه با مفهوم ربط را عنوان می‌کند. بر این مبنا می‌توان سنجه‌های ارزیابی کارآمدی موتورهای کاوش را با استفاده از رویکردهای موجود در بازیابی اطلاعات تبیین نمود. در این راستا بر اساس نظر سیرتکین (۲۰۱۲) دو سنجه متوسط فاصله و بی پرف مبتنی بر رویکرد نظام گرایی هستند؛ درحالی‌که سنجه سود تجمعی تعدیل‌یافته نرمال به کاربران اهمیت بیشتری می‌دهد. افزون بر این، بورلاند و اینگورسن (۱۹۹۸) ادعا می‌کنند که سنجه اتحاد جاکارد و اتحاد کسینوسی آن‌ها ریشه در رویکرد سوم و دوگان‌گرا دارد اما به نظر می‌رسد که میزان توجه به هر دو رویکرد کم‌تر است؛ زیرا در اصل در سنجه آن‌ها نیز میزان توافق ربط موتور کاوش و ربط اختصاص داده‌شده به مدارک از سوی کاربران محاسبه می‌شود. در این رابطه شاید بتوان سنجه‌های دقت، بازیافت، جامعیت را ازجمله سنجه‌های غیر رتبه‌بندی دسته‌بندی نمود که متأثر از رویکرد نظام‌گرایی هستند؛ درحالی‌که سنجه‌های اتحاد جاکارد، اتحاد کسینوسی، متوسط فاصله، سود تجمعی تعدیل‌یافته نرمال، بی پرف و سنجه رتبه کارآمدی را در گروه سنجه‌های رتبه‌بندی قرارداد که متأثر از رویکرد کاربرگرایی هستند. افزون بر دو رویکرد یادشده، رویکرد سومی نیز در پژوهش‌های بازیابی اطلاعات نظیر ثورنلی و گیب[1] (۲۰۰۷) مطرح‌شده است که در این رویکرد به تعامل و یگانش دو رویکرد موجود تأکید می‌شود.

بر اساس نظر مشهور پژوهشگران بازیابی اطلاعات نظیر میزاور (۲۰۰۱)، کرافت، متزلر و استروهمن[2] (۲۰۱۵)، حریری و همکاران (۱۳۹۳) در محاسبه سنجه‌های دقت و بازیافت، مقیاس قضاوت ربط دودویی است. به‌بیان‌دیگر مدارک یا مرتبط هستند یا نامرتبط؛ درحالی‌که این امر چندان اعتبار ندارد؛ بلکه می‌توان با استفاده از قضاوت ربط پیوسته نیز سنجه‌های دقت و بازیافت را محاسبه نمود. برای مثال می‌توان به پژوهش کومار و پراکاش (۲۰۰۹)، عباسی دشتکی و چشمه سهرابی (۱۳۹۸) و زینالی تازه‌کندی و نوکاریزی (۲۰۲۰) اشاره نمود که از مقیاس قضاوت ربط پیوسته استفاده و درعین‌حال دو سنجه دقت و بازیافت را نیز محاسبه کرده‌اند.

همان‌گونه که اینگورسن (۱۳۸۹) نیز تأکید می‌کند، هرکدام از سنجه‌های مطرح‌شده از یک مدل هنجاری تبعیت می‌کند. بر این مبنا می‌توان گفت که مدل هنجاری سنجه دقت، بازیابی نشدن مدارک نامرتبط (کرافت و ...، ۲۰۱۵)؛ سنجه بازیافت، بازیابی مدارک مرتبط یا با نمره بالای موجود در پایگاه نمایه موتور کاوش (کلارک و ویلت،۱۹۹۷)؛ سنجه جامعیت، بازیابی مدارک با نمره ربط بالای موجود در وب است (نوکاریزی و زینالی تازه کندی، ۲۰۱۹) و مدل هنجاری سنجه‌های متوسط فاصله، سود تجمعی تعدیل‌یافته نرمال، بی پرف و رتبه کارآمدی بازیابی مدارک با نمره ربط بالا قبل از مدارک با نمره ربط پایین است (سیرتکین، ۲۰۱۲). همچنین مدل هنجاری سنجه اتحاد جاکارد و اتحاد کسینوسی، شباهت و نزدیکی نمره ربط نظام به نمره ربط معیار است (بورلند، ۲۰۰۳).

در رابطه با تأثیرپذیری سنجه‌های یادشده می‌توان گفت که سنجه دقت به‌اندازه پایگاه موتور کاوش بستگی دارد، درحالی‌که سنجه بازیافت از اندازه پایگاه موتور کاوش هیچ تأثیری نمی‌پذیرد. از طرف دیگر سنجه جامعیت کاملاً وابسته به‌اندازه پایگاه موتور کاوش است. همچنین هر سه سنجه یادشده از این‌که مدارک مرتبط‌تر جلوتر از مدارک مرتبط بازیابی شوند، تأثیر نمی‌پذیرند. از سویی دیگر سنجه‌های گروه رتبه‌بندی که قبلاً ذکر شد بر این مبنا طراحی‌شده‌اند که باید مدارک مرتبط‌تر، جلوتر از مدارک مرتبط بازیابی شوند، اما این سنجه‌ها نیز به این مورد که در کل مدارک مرتبط‌تر بازیابی شوند نه مدارک مرتبط اهمیت چندانی نمی‌دهند. همچنین انتظار می‌رود که سنجه‌های اتحاد جاکارد و اتحاد کسینوسی به رتبه‌بندی مدارک حساس باشند؛ درحالی‌که این انتظار چندان برآورد نشده است و یگانش دو رویکرد ربط نظام‌گرا و ربط کاربرگرا در این دو سنجه یادشده باکمی تردید همراه است.

در پیوند بین معماری موتورهای کاوش و سنجه‌های ارزیابی موتورهای کاوش می‌توان گفت که در سنجه بازیافت عملکرد مؤلفه گردآوری منابع و مؤلفه رتبه‌بندی موردسنجش قرار نمی‌گیرد؛ درحالی‌که در همه سنجه‌های رتبه‌بندی به عملکرد مؤلفه رتبه‌بندی مدارک تأکید شده است. همچنین به نظر می‌رسد که هیچ‌کدام از سنجه‌های بررسی‌شده در پژوهش به عملکرد مؤلفه گردآوری منابع توجه نشده است؛ درحالی‌که چرایی طراحی سنجه جامعیت تأکید به اهمیت عملکرد مؤلفه گردآوری منابع بوده است.

❖ مرور حاضر به‌منظور تحلیل سنجه‌های ارزیابی ربط موتورهای کاوش در توجه به کاربرد هر یک و نقص‌های آن‌ها در مقایسه با یکدیگر انجام گرفت. تعیین کارآمدی آن‌ها برای موتورهای کاوش، منوط به انجام پژوهش‌های آتی است. در این خصوص به پژوهشگران ارزیابی موتورهای کاوش پیشنهاد

[1] Thornley and Gibb
[2] Croft, Metzler and Strohman

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

❖ می‌شود که در صورت امکان، سنجه‌هایی را از هر دو گروه غیر رتبه‌بندی و رتبه‌بندی جهت تصمیم در مورد کارآمدی موتورهای کاوش انتخاب نمایند. درنهایت با توجه به پژوهش‌های کلارک و ویلت، (۱۹۹۷) و پراکاش و کومار (۲۰۰۹) و نوکاریزی و زینالی تازه کندی (۲۰۱۹) از بین سنجه‌های گروه غیر رتبه‌بندی، سنجه جامعیت و با توجه به پژوهش‌های ساکی و کاندو (۲۰۰۸) و سیرتکین (۲۰۱۲) از بین سنجه‌های رتبه‌بندی، سنجه سود تجمعی تعدیل‌یافته نرمال سنجه‌های قوی‌تری از سنجه‌های مشابه خود هستند. همچنین ازآنجایی‌که در متون روش پژوهش نیز از روش ترکیبی یادشده است که در این رویکرد باور بر این است که کاستی‌های روش کمی و کیفی برطرف شده است. بر این مبنا پیشنهاد می‌شود که پژوهشی در قالب پایان‌نامه یا طرح پژوهشی به طراحی سنجه‌ای برای ارزیابی موتورهای کاوش پرداخته شود تا با استفاده از این سنجه بتوان عملکرد همه مؤلفه‌های موتورهای کاوش یعنی مؤلفه گردآوری منابع، مؤلفه تبدیل شکل منابع، مؤلفه نمایه‌سازی، مؤلفه بازیابی و مؤلفه رتبه‌بندی را به‌صورت یکجا و کل واحد موردسنجش قرارداد؛ چراکه هرکدام از سنجه‌های موجود یک یا دو مؤلفه از مؤلفه‌های موتور کاوش را موردتوجه قرار داده‌اند. افزون بر این، در این پژوهش ادعا نمی‌شود که همه سنجه‌های مهم ارزیابی موتورهای کاوش مرور شده است؛ بلکه نیاز به پژوهش‌های بیشتری در این زمینه وجود دارد تا با توصیف و تحلیل سنجه‌های یادشده، نقاط قوت و ضعف آن‌ها آشکار شود تا بدین‌وسیله پژوهش‌های قوی‌تری درزمینهٔ ارزیابی موتورهای کاوش انجام شود.

تقدیر و تشکر

بدین وسیله از کلیه افرادی که در انجام پژوهش حاضر همکاری نمودند، تشکر و قدردانی به عمل می‌آید.

تعارض منافع

نویسندگان، اعلام می‌دارند در رابطه با انتشار مقاله ارائه‌شده، هیچ‌گونه تعارض منافعی وجود ندارد.

منبع حمایت‌کننده

پژوهش حاضر، پژوهشی مستقل و بدون دریافت هرگونه حمایتی انجام شده است.

References

Abbasi Dashtaki, N; Cheshmeh Sohrabi, M (2019). Google, Yahoo and Bing Search Engines' Performance in the Persian Information Retrieval: A Fuzzy and Classical Evaluation. Journal of National Studies on Librarianship and Information Organization ,30(2),96-111. (Persian)

Al-Maskari, A., Sanderson, M., & Clough, P. (2007, July). The relationship between IR effectiveness measures and user satisfaction. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 773-774).

Asadi Qadikolaei O., Asadi S., Noroozi A., Ehsani R, (2014). A Comparison of Precision in General Search Engines and Specialized Databases for Radiology Image Retrieval, Journal of Health Management, 5(2), 77-87. (Persian)

Azadi, Gh (2005). The scale of web search engines precision in information retrieval of library and information science discipline, National Studies on Librarianship and Informaion Organization, 16(3), 111. (Persian)

Babaei, E and Sajedi, M (2013). A Comparative Study on Efficiency of Medical Specialized Search Engines in Retrieving Information Related to Gynecology and Obstetrics, Health Information Management, 10(2), 234. (Persian)

Bama, S. S., Ahmed, M. I., & Saravanan, A. (2015). A survey on performance evaluation measures for information retrieval system. International Research Journal of Engineering and Technology, 2(2), 1015-1020.

Bar-Ilan, J. (1998). On the overlap, the precision and estimated recall of search engines. A case study of the query "Erdos". Scientometrics, 42(2), 207-228.

Bayanvand, A (2012). The Basics of Computer and Internet.Tehran: Chapar.(Persian)

Bilal, D. (2012). Ranking, relevance judgment, and precision of information retrieval on children's queries: Evaluation of Google, Y ahoo!, B ing, Y ahoo! K ids, and ask K ids. Journal of the American Society for Information Science and Technology, 63(9), 1879-1896.

Borlund, P. & Ingwersen, P. (1998) Measures of relative relevance and ranked half-life: performance indicators for interactive IR. In: Croft, B.W, Moffat, A., van Rijsbergen, C.J., Wilkinson, R., and Zobel, J., eds.

Borlund, P (2003). The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. Information Research, 8(3). [Available at: http://informationr.net/ir/8-3/paper152.html]

Buckley, C., & Voorhees, E. M. (2004, July). Retrieval evaluation with incomplete information. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 25-32). ACM.

Budd, J M (2004). Relevance: Language, semantics, philosophy. Library trend, 52(3).

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]

[DOR: 20.1001.1.24237418.1399.7.2.2.6]

Büttcher, S., Clarke, C. L., & Cormack, G. V. (2016). Information retrieval: Implementing and evaluating search engines. Mit Press.

Clarke, S. J., & Willett, P. (1997). Estimating the recall performance of Web search engines. Aslib Proceedings, 49 (7), 184-189.

Comperhensive list of search engines (2017), In The search engine list website, Retrivited 9 Dec. 2017 from http://www.thesearchenginelist.com/

Croft, W. B., Metzler, D., & Strohman, T. (2015). Search engines: Information retrieval in Practice. London: Pearson Education.

Daverpanah, MR (2004). Paradigm and information retrieval. Iranian Journal of Library and Information Science, 7(3), 2-14.(Persian)

Mea, D; Gianluca,V; Luca, D; Gaspero, D and Mizzaro, S. (2006) Measuring retrieval effectiveness with average distance measure (ADM). Information Wissenschaft und Praxis 57(8), 433-443.

Demirci, R. G., Kismir, V. and Bitirim, Y. (2007), An evaluation of popular search engines on finding turkish document, Second International Conference on Internet and Web Applications and Services (ICIW'07), IEEE, Turkey, pp.1-5.

Thornley, C. and Gibb, F. (2007). A dialectical approach to information retrieval. Journal of documentation, 63 (5), 755-764.

Fidel, R. (2008). Are we there yet? Mixed methods research in library and information science. Library & Information Science Research, 30(4), 265-272.

Frické, M. (1998), "Measuring recall", Journal of information science,24 (6), 409-417.

Garoufallou, E. (2012). Evaluating search engines: A comparative study between international and Greek SE by Greek librarians. Program: electronic library & information systems, 46(2), 182-198.

Ghiasi, M; Daliri, S; Kouchakinejad, L and Abbasian Joushghani, A (2015). A Comparison of Accuracy in Specialized Medical Search and General Search Engines for Retrieving Medical Image, Educational Developement of Jundishapur, 6(2), 131-138. (Persian)

Goel, S., & Yadav, S. (2012). An Overview of Search Engine Evaluation Strategies. International Journal of Applied Information Systems, 1, 7-10.

Hariri N, Babalhavaeji F, Farzandipour M, Nadi Ravandi S(2014). Evaluation Criteria of Information Retrieval Systems: What We Know and What We Do Not Know.; Iranian Research Institute for Information Science and Technology, 30 (1):199-221.(Persian)

Hariri, N and Vakili Mofrad, H (2014). A Comparison of the Precision of General and Specialized Medical Search Engines in Medical Images Retrieval, Health Information Management, 10(6), 830-839. (Persian)

Hariri, N. (2011). Relevance ranking on Google. Online Information Review. 35(4), 598-610.

Henzinger, M. (2007). Search technologies for the Internet. Science, 317(5837), 468-471.

Hjørland, B. (2010). The foundation of the concept of relevance. Journal of the American Society for Information Science and Technology, 61 (2), 217-237.

Ingwersen, P. (2010). Information retrieval interaction. Translated by Hajar Setodeh.Tehran: Ketabdar.

Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems (TOIS), 20(4), 422-446.

Kent, A., Berry, M. M., Luehrs Jr, F. U., and Perry, J. W. (1955), "Operational criteria for designing information retrieval systems", American documentation, 6 (2), 93-101.

Kosha, K (2002). Internet Exploration Tools: Search Principles, Skills, and Features. Tehran: Ketabdar. (Persian).

Kumar, B. S., & 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(1), 124-13

Kumar, B. T., and Sampath Pavithra, S. M. (2010), Evaluating the searching capabilities of search engines and metasearch engines: a comparative study, Annals of Library and Information Studies), 57 (2), 87-97.

Kumar, K. and Bhadu, V. (2013), A comparative study of BYG search engines, American Journal of engineering research, 2 (4), 39-43.

Lancaster, F. W. (1979), Information retrieval systems; characteristics, testing, and evaluation (2nd ed ed.), Wiley, New York..

Landoni, M. and Bell, S. (2000), Information retrieval techniques for evaluating search engines: a critical overview, Aslib Proceedings, 52 (3), 124-129.

Lewandowski, D. (Ed.). (2012). Web search engine research. Emerald Group Publishing Limited.

Lopes, C. T., and Ribeiro, C. (2011). Comparative evaluation of web search engines in health information retrieval. Online Information Review, 35(6), 869-892.

Mea, V. D., & Mizzaro, S. (2004). Measuring retrieval effectiveness: A new proposal and a first experimental validation. Journal of the American Society for Information Science and Technology, 55(6), 530-543.

Mizzaro, S. (2001, September). A new measure of retrieval effectiveness (or: What's wrong with precision and recall). In International workshop on information retrieval (IR'2001) (pp. 43-52). Infotech Oulu.

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]
[DOR: 20.1001.1.24237418.1399.7.2.2.6]

Mohammad Esmaeil, S and Mansour Kiaie, R (2012). A Comparison between Search Engines and Meta-search Engines in Retrieving Information Related to Physics and the Extent of their Overlap, National Studies on Librarianship and Informaion Organization, 22(3), 130. (Persian)

Mohammadesmaeil, S; Lafzighazi, E and Gilvari, A (2008). Comparing Search Engines and Meta-search Engines in Pharmaceutics Information Retrieval, Health Information Management, 5(2), 121-129. (Persian)

Mohammadesmeil, S and Naraghian, N (2017). Comparing Search Engines and Meta Search Engines in Dentistry Information Retrieval, Journal of Research in Dental Sciences, 14(2), 118-127. (Persian)

Nowkarizi, M; Zeynali Tazehkandi, M. (2019). Rethinking the Recall Measure in Appraising Information Retrieval Systems and Providing a New Measure by Using Persian Search Engines. International Journal of Information Science and Management,17(1), 1-16.

Nowkarizi,M and Zeynali Tazehkandi, M (2017). The overlap and coverage of 4 local search engines: Parsijoo, Yooz, Parseek and Rismoun, Human Information Interaction, 4(3), 48-59. (Persian)

Pao, M. L. (2000). Concepts of information retrieval. Translated by Asad Olah Azad and RahmattolahFattahi.Mashhad: Ferdowsi university of Mashhad. (Persian)

Powers, D.M.W., 2011. Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. Journal of Machine LearningTechnologies, 2(1), 37-63.

Rajabi,M and Norouzi, Y (2015). Persian Search Engines: Evaluating Search Features, Information Retrieval, Precision and Recall and Their Overlaps, National Studies on Librarianship and Informaion Organization, 26(3), 133-150. (Persian)

Riahinia N, Rahimi F and AllahBakhshian L (2015). Matching Scores of System Relevance and User-Oriented Relevance in SID, ISC and Google Scholar. Human Information Interaction, 2 (1), 1-11. (Persian)

Sakai, T. (2007, July). Alternatives to bpref. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 71-78). ACM.

Sakai, T. (2012, April). Evaluation with informational and navigational intents. In Proceedings of the 21st international conference on World Wide Web (pp. 499-508).

Sakai, T., & Kando, N. (2008). On information retrieval metrics designed for evaluation with incomplete relevance assessments. Information Retrieval, 11(5), 447-470.

Sakai, T., & Song, R. (2011, July). Evaluating diversified search results using per-intent graded relevance. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 1043-1052).

Saracevic, T. (2007). Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science: Nature and Manifestations of Relevance. Journal of the American Society for Information Science and Technology, 58 (13), 1915–1933.

Saracevic, T. (2015). Why is relevance still the basic notion in information science. In Re: inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015) (pp. 26-36).

Shafi, S. and Rather, R. A. (2005), "Precision and recall of five search engines for retrieval of scholarly information in the field of biotechnology", Webology, 2 (2), 42-47.

Shang, Y., and Li, L. (2002). Precision evaluation of search engines. World Wide Web, 5(2), 159-173.

Sirotkin, P. (2012). On Search Engine Evaluation Metrics. arXiv preprint arXiv:1302.2318.

Smith, A. G. (2003). Think local, search global? Comparing search engines for searching geographically specific information. Online Information Review., 27(2), 102-109.

Soleymani, H (2009).Web search and database training. Tehran: Hojatolah Soleymani

Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. Information Processing & Management, 40(4), 677-691.

Yilmaz, E., Carterette, B., & Kanoulas, E (2012). Evaluating Web Retrieval Effectiveness. In Dirk lewandowski, web search engine research. Bingley, west Yorkshire: Emerald Group Publishing Limited.

Yosefi, A (1997). False drop in information storage and retrieval. Iranian Journal of Scientific Information and Documentation Center, 13(1), 1-9.(Persian).

Zeynali Tazehkandi, M. and Nowkarizi, M. (2020). Evaluating the effectiveness of Google, Parsijoo, Rismoon, and Yooz to retrieve Persian documents. Library Hi Tech. https://doi.org/10.1108/LHT-11-2019-0229.

Zhou, B., & Yao, Y. (2010). Evaluating information retrieval system performance based on user preference. Journal of Intelligent Information Systems, 34(3), 227-248.Zuva, K. and Zuva, T. (2012), "Evaluation of information retrieval systems", International journal of computer science and information technology, 4 (3),35-43.

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]

[DOR: 20.1001.1.24237418.1399.7.2.2.6]

Croft, W. B., Metzler, D., & Strohman, T. (2015). Search engines. In Information Retrieval in Practice. Pearson Education, Inc.

Zuva, K., and Zuva, T. (2012). Evaluation of information retrieval systems. International journal of computer science & information technology, 4(3), 35-43.

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]

[DOR: 20.1001.1.24237418.1399.7.2.2.6]

[Downloaded from c4i2016.khu.ac.ir on 2026-11-05]

[DOR: 20.1001.1.24237418.1399.7.2.2.6]