Archive of SID

# Article Ranking by Recommender Systems vs. Users' Perspectives

S. Sadein Khorram[1]| J. Abbaspour[2]

Received: 26, May 2019
Accepted: 31, July 2019

**Purpose**: Compare the rankings of articles by Google Scholar and Web of Science recommender systems against users' perspectives.

**Methodology**: 120 PhD candidates of Shiraz University in the fields of humanities, sciences, engineering and agriculture, (30 from each field) voluntarily participated in the study. They were asked to introduce three articles had recently read for their thesis. One of the three which was indexed by both databases was chosen and named as the core article. For each core article 10 recommended articles recommended by each recommender system were retrieved (2,400 overall). Using imitating software exclusively designed for this study, participants were asked to rank the articles retrieved by the two recommender systems for their core articles. Normalized Discounted Cumulative Gain (NDCG) measure was employed for analysis.

**Findings**: There was a noticeable but weak relationship between the users' assigned rankings and the rankings of Google Scholar and Web of Science databases. Correlation between the rankings of both databases with NDCG measure was also weak.

**Conclusion**: The algorithms used for ranking by both recommender systems hardly in matched that of the users. Therefore, ranking algorithms of both databases may need some revision

**Keywords:**

Documents' ranking, Research paper recommender systems, NDCG measure, Information retrieval, Google Scholar, Web of Science

DOI: 10.30484/nastinfo.2019.2187.1838

1. M.A, Knowledge and Information Science, Shiraz University, sabasadein@gmail.com
2. Assistant Professor, Knowledge and Information Science, Shiraz University (Corresponding author), javad.abbaspour@gmail.com

www.SID.ir

Archive of SID

مطالعات ملی کتابداری و سازماندهی اطلاعات
NASTINFO

# سنجش رتبه‌بندی سامانه‌های پیشنهاددهندهٔ مقاله در تقابل با رتبه‌بندی کاربران

صبا سعدین‌خرم[1] | جواد عباس‌پور[2]

**هدف:** سامانه‌های پیشنهاددهنده مقاله‌های علمی در تلاش‌اند به‌جای الزام کاربر به اصلاح راهبرد جستجو و فرمول‌بندی واژگان پرسش، الگوریتم‌ها و رویکردهای مناسب را برای عرضه مرتبط‌ترین مدارک به‌کار گیرند. علاوه بر ربط، رتبه‌بندی مدارک بازیابی‌شده نیز می‌تواند رضایت کاربران را تضمین کند. این مقاله نتیجه تحلیل رتبه مقاله‌ها را در سامانه‌های پیشنهاددهنده پایگاه وب‌آوساینس و موتور جستجوی گوگل‌اسکالر از دیدگاه کاربران و سنجه Normalized Discounted Cumulative Gain (NDCG) گزارش می‌کند.

**روش‌شناسی:** از ۱۲۰ دانشجوی داوطلب دکترای دانشگاه شیراز در رشته‌های علوم انسانی، علوم پایه، فنی-مهندسی، کشاورزی، و دام‌پزشکی (از هریک ۳۰ نفر) خواسته شد ۲۴۰۰ مقاله (۱۲۰۰ مقاله پیشنهادی گوگل‌اسکالر و ۱۲۰۰ مقاله پایگاه وب‌آوساینس) را به‌لحاظ ربط، رتبه‌بندی کنند. داده‌ها با پرسشنامه و نرم‌افزار پژوهشگرساخته گردآوری شد.

**یافته‌ها:** میان رتبه انتسابی کاربران و رتبه انتسابی پایگاه شباهت ضعیف بود. شباهت رتبه‌بندی مقالات در سامانه‌های هر دو پایگاه نیز ضعیف، اما معنادار بود. الگوریتم‌ها و شاخص‌های سامانه‌های پیشنهاددهنده هر دو پایگاه برای رتبه‌بندی چندان موفق نیست و نیاز به بازنگری دارد.

**نتیجه‌گیری:** الگوریتم‌ها و شاخص‌هایی که سامانه پیشنهاددهنده دو پایگاه برای رتبه‌بندی مقاله‌های مرتبط پیشنهادی درنظر گرفته‌اند، در تأمین رضایت کاربران زیاد موفق نبوده است؛ بنابراین بازنگری در الگوریتم‌های رتبه‌بندی این دو پایگاه ضروری به‌نظر می‌رسد.

دریافت: ۹۸/۰۳/۰۶ | پذیرش: ۹۸/۰۵/۱۰

۱. کارشناس ارشد علم اطلاعات و دانش‌شناسی، دانشگاه شیراز sabasadein@gmail.com
۲. استادیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه شیراز (نویسنده مسئول) javad.abbaspour@gmail.com

**کلیدواژه‌ها**
رتبه‌بندی مدارک، سامانه‌های پیشنهاددهنده مقالات، سنجه NDCG، بازیابی اطلاعات، پایگاه وب‌آوساینس، موتور جستجوی گوگل‌اسکالر

*مطالعات ملی کتابداری و سازماندهی اطلاعات*، دوره سی، شماره سوم، پاییز ۱۳۹۸، ص ۴۶-۵۷.

DOI: 10.30484/nastinfo.2019.2187.1838

www.SID.ir

Archive of SID



**مقدمه**

پژوهش نشان داده است از میان گاه صدها مدرکی که سامانه‌های پیشنهاددهنده عرضه می‌کنند کاربران فقط به ۱۰ مدرک نخست (Drori,. 2002; Lewandowski, 2008; Nowicki, 2003; Bar-Ilan, Levene, Mat-Hassan, 2006) و حتی گاه فقط به سه نتیجه نخست توجه می‌کنند (Enquiro, 2005, quoted in Bar-Ilan, Keenoy, Levene, & Yaari, 2009). از این‌رو، رتبه‌بندی مقالات برای جلب رضایت کاربران و موفقیت سامانه بسیار اهمیت دارد.

چندین پژوهشگر رتبه‌بندی سامانه را از دیدگاه کاربر ارزیابی کردند (ره‌گشای، 1390؛ Su, 2003; Nowicki, 2003; Beg, 2005; Patil, Alpert, Karat, & Wolf, 2005; Bar-Ilan, Keenoy, Yaari, & Levene, 2007; Hariri, 2011). سنجه NDCG[1] روشی جدید برای ارزیابی نتایج بازیابی براساس شاخص اثربخشی، یعنی دقت[2] از نگاه استفاده‌کننده است؛ با این تفاوت که با استفاده از آن می‌توان رتبه‌بندی را براساس ربط چنددرجه‌ای ارزیابی کرد نه همچون گذشته دودرجه‌ای (مرتبط/نامرتبط). ربط چنددرجه پیشتر برای بررسی اثربخشی سامانه‌های بازیابی اطلاعات به‌کار رفته است (Teevan, Dumais, & Horvitz, 2005; Kekäläinen, 2005; Sakai, 2007; Al-Maskari, Sanderson, & Clough, 2007; Sanderson, Paramita, Clough, & Kanoulas, 2010).

الگوریتم‌های سامانه‌های پیشنهاد مدارک به‌دلایل تجاری و رقابت شرکت‌ها طراحی شده‌اند؛ از این‌رو، نتایج آنها یکسان نیست. الگوریتم‌های آنها متن‌محور یا پیوندمحور است. در الگوریتم‌های متن‌محور شباهت تعداد اصطلاحات مشترک بین پرسش و مدرک یا دو مدرک اندازه‌گیری می‌شود (Yoon, Kim, Kim, & Hwang, 2011) و شباهت با بیشترشدن شمار واژه‌های مشترک بیشتر می‌شود. بازیابی و رتبه‌بندی متن‌محور به‌سبب مسائل مربوط به پردازش زبان طبیعی، همچون نامشخص‌بودن مرجع ضمیرها، چندمعنایی، و هم‌معنی‌بودن برخی واژه‌ها دشوار است (Vallez & Pedraza-Jimenez, 2007). سنجه‌های فراوانی واژه- معکوس فراوانی مدرک (تی‌اف-ای دی‌اف)[3] و شباهت کسینوسی از سنجه‌های متن‌محور است. سنجه‌های استنادمحور پیوندهای استنادی را مبنای شباهت مدارک می‌گیرند و می‌توانند مدارکی را بیابند که نمایه‌سازی کلیدواژه‌ای آنها را نشناخته است (Pao,1993; Smith, 1981, quoted in Eto, 2013). سنجه‌های زوج کتاب‌شناختی، هم‌استنادی، و استنادهای مشترک در معکوس فراوانی در این دسته جای می‌گیرند. به‌کارگیری الگوریتم‌ها و فنون پیچیده رتبه‌بندی به کاربران کمک می‌کند مرتبط‌ترین اطلاعات را در کمترین زمان و با ساده‌ترین روش بیابند (Bar-Ilan et al., 2009). از این‌رو، ارزیابی سامانه‌های بازیابی اطلاعات برای مشخص‌شدن میزان موفقیت‌شان در تأمین

1. Normalized Discounted Cumulative Gain
2. Precision
3. Term Frequency-Inverse Document Frequency (TF-IDF)



www.SID.ir

Archive of SID

رضایت کاربران ضروری است.

نظر به اهمیت و جامعیت پایگاه‌های وب‌آوساینس و موتور جستجوی گوگل‌اسکالر، پژوهش حاضر رتبه‌بندی مقاله‌ها را در سامانه‌های پیشنهاددهنده این دو پایگاه از دیدگاه کاربران و سنجه NDCG ارزیابی می‌کند.

## روش‌شناسی

۱۲۰ نفر از دانشجویان دکترای دانشگاه شیراز در پنج گروه علوم انسانی، علوم پایه، فنی–مهندسی، کشاورزی، و دام‌پزشکی داوطلبانه در این پژوهش شرکت کردند. آنها با این فرض انتخاب شدند که بنا به ماهیت پژوهشی مقطع تحصیلی و دانش زبان انگلیسی‌شان با این دو پایگاه بیشتر آشنا هستند. در هر گروه ۳۰ نفر شرکت داشتند. این تعداد را برای مطالعات مقایسه‌ای کافی دانسته‌اند (دلاور، ۱۳۹۶).

ابتدا از شرکت‌کنندگان خواسته شد سه مقاله انگلیسی را معرفی کنند که اخیراً مطالعه کرده بودند. ما یکی از سه مقاله را که در هر دو پایگاه موجود بود به‌عنوان مقاله هسته انتخاب و برای هرکدام از آنها ۱۰ مقاله نخست سامانه‌های دو پایگاه را با دادن دستور استخراج «مقاله‌های مرتبط»[1] بازیابی کردیم. به‌ازای هر مقاله هسته ۱۰ مقاله صفحه نخست پیشنهادشده از هریک از دو پایگاه را در اختیار شرکت‌کنندگان نهادیم. سبب اکتفا به ۱۰ مقاله صفحه نخست آن بود که پژوهش نشان داده است بیش از نیمی از کاربران (۵۸ درصد) فقط به صفحه اول نتایج بازیابی‌شده توجه می‌کنند (Jansen, Spink, & Saracevic, 2000). در پژوهش‌هایی دیگر نیز تعداد ۱۰ نتیجه اول بازیابی‌شده از سامانهٔ بازیابی اطلاعات مبنای کار پژوهشگران قرار گرفته است (Kinley, Tjondronegoro, Partridge, & Edwards, 2014؛ فرهودی و حریری، ۱۳۹۲؛ Joachims, Granka, Pan, Hembrooke, & Gay, 2005؛ ریاحی‌نیا، رحیمی، لطیفی، و بخشیان، ۱۳۹۴).

اطلاعات کتاب‌شناختی ۱۰ مقاله در یک فایل اکسل ذخیره و برای هریک دو فایل اکسل باز کردیم؛ یکی شامل کلیدواژه‌ها، عنوان‌ها، و چکیده‌های ۱۰ مقاله ابتدایی بازیابی‌شده در گزینه مقاله‌های مرتبط پیشنهادی موتور جستجوی گوگل‌اسکالر و دیگری شامل مقاله‌های مرتبط پیشنهادی پایگاه وب‌آوساینس بود. تقدم و تأخر مقاله‌های این فهرست به‌طور تصادفی تغییر می‌کرد. علت این بود که رتبه‌بندی پیش‌فرض مقاله‌ها در سامانه بازیابی اطلاعات بر قضاوت کاربر تأثیر نگذارد و این عامل کنترل شود.

1. Related articles


www.SID.ir

Archive of SID



سپس از شرکت‌کنندگان خواستیم با استفاده از نرم‌افزار شبیه‌ساز، رتبه‌بندی مقالات را انجام دهند. این نرم‌افزار را یک متخصص برنامه‌نویسی طراحی کرده و در چند آزمایش مشکلات آن برطرف شده بود و سه استاد رشته علم اطلاعات و دانش‌شناسی روایی آن را تأیید کرده بودند.

فایلِ اکسلِ مقاله‌های پیشنهادشده در محیط نرم‌افزار بارگذاری شد. کاربران با انتخاب گزینه «شروع» به صفحه‌ای هدایت می‌شدند و هریک از مقالات را با دادن امتیاز ۱ تا ۱۰ از نظر میزان ربط با مقاله هسته رتبه‌بندی کردند. مقالات هر پایگاه جداگانه رتبه‌بندی شد. پایگاه نخست را نرم‌افزار به‌طور تصادفی برای کاربر انتخاب می‌کرد تا احتمال درست انجام‌نشدن رتبه‌بندی مقاله‌های پایگاه دوم به‌سبب خستگی یا کمبود وقت برطرف شود. برای تعیین همبستگی میان متغیرها از آزمون ناپارامتری اسپیرمن استفاده شد. همبستگی‌ها با استفاده از نرم‌افزار اس‌پی‌اس‌اس نسخه ۲۵ محاسبه شد.

برای محاسبه NDCG کاربران با استفاده از نرم‌افزار شبیه‌ساز براساس مقیاسِ سه‌درجه‌ای شامل «مرتبط، تاحدی مرتبط، و نامرتبط» درباره ۱۰ مقاله هریک از دو پایگاه با مقاله هسته خود داوری کردند. امتیازهای آنها مبنای محاسبه با سنجه NDCG شد که درجه ربط را نشان می‌دهد و به‌خاطر نرمال‌کردن اعداد، مقایسه عملکرد سامانه‌ها با آن ممکن است. برای محاسبه NDCG ابتدا با استفاده از فرمول زیر مقدار DCG[1] برای مدارک موجود در هر فایل محاسبه شد.

$$DCG_n = rel_1 + \sum_{i=2}^{n} \frac{rel_i}{\log_2 i}$$

برای نرمال‌کردن مقدار DCG، مدارک باید براساس نظم ایده‌آل و در بهترین حالت ممکن رتبه‌بندی شوند. در این صورت، به آن IDCG[2] گویند. DCG نرمال طبق فرمول زیر از تقسیم DCG بر IDCG به‌دست می‌آید و مقداری از صفر تا یک می‌پذیرد. برای هر مدرک مقدار DCG و IDCG و سپس NDCG محاسبه شد (Wang, Li, He, & Liu, 2013).

$$NDCG_n = \frac{DCG_n}{IDCG_n}$$

1. Discounted Cumulative Gain
2. Ideal Discounted Cumulative Gain



www.SID.ir

Archive of SID

برای بررسی رابطه میان رتبه‌بندی مقاله‌های مرتبطِ پیشنهادی از دیدگاه کاربران و رتبه‌بندی موتور جستجوی گوگل‌اسکالر، به‌دلیل رتبه‌ای‌بودن متغیرها از آزمون همبستگی اسپیرمن استفاده شد.

**یافته‌ها**

یافته‌های حاصل از ضریب همبستگی در جدول ۱ گزارش شده است. نتایج نشان می‌دهد بین این دو متغیر به‌لحاظ آماری همبستگی وجود دارد. با درنظرگرفتن مقدار ضریب همبستگی به‌دست‌آمده (rs=0/151, P<0/01) و با توجه به مثبت‌بودن این عدد، می‌توان گفت میان این دو رتبه‌بندی به‌لحاظ آماری رابطه‌ای مستقیم و معنادار وجود دارد؛ اما رابطه به‌سبب کوچک‌بودن ضریب همبستگی ضعیف است. بنابراین، در پایگاه گوگل‌اسکالر رتبه‌بندی پایگاه از مقاله‌ها با رتبه‌بندی کاربران شباهت دارد؛ ولی درخور توجه نیست.

**جدول ۱.** نتایج آزمون همبستگی رتبه‌بندی مقاله‌های مرتبط پیشنهادی پایگاه گوگل‌اسکالر با رتبه‌بندی کاربران

| | رتبه‌بندی کاربران | رتبه‌بندی گوگل‌اسکالر |
|---|---|---|
| رتبه‌بندی گوگل‌اسکالر | ۰/۱۵۱* | ۱ |
| رتبه‌بندی کاربران | ۱ | ۰/۱۵۱* |

*P<0/01, N =120

برای تعیین رابطه میان رتبه‌بندی دیدگاه کاربران و رتبه‌بندی پایگاه وب‌آوساینس از آزمون همبستگی اسپیرمن

استفاده شد. بین این دو متغیر همبستگی ضعیف (rs=0/151, P<0/01) وجود داشت (جدول ۲).

**جدول ۲.** نتایج آزمون همبستگی رتبه‌بندی مقاله‌های مرتبط پیشنهادی پایگاه وب‌آوساینس

| | رتبه‌بندی کاربران | رتبه‌بندی وب‌آوساینس |
|---|---|---|
| رتبه‌بندی وب‌آوساینس | ۰/۱۱۷* | ۱ |
| رتبه‌بندی کاربران | ۱ | ۰/۱۱۷* |

*P<0/01, N =120


www.SID.ir

Archive of SID



برای مقایسه رتبه‌بندی مقاله‌های مرتبط پیشنهادی پایگاه وب‌آوساینس و موتور جستجوی گوگل‌اسکالر براساس سنجه NDCG از آزمون ناپارامتری ویلکاکسون استفاده شد. همان‌طور که جدول ۳ نشان می‌دهد با توجه به z=-0/699 و P>0/05، براساس سنجه NDCG تفاوت معنادار بین رتبه‌بندی مقاله‌های مرتبط پیشنهادی این دو پایگاه وجود ندارد. پس دو پایگاه از حیث نحوه رتبه‌بندی مقاله‌های مرتبط پیشنهادی و نحوه عملکرد باهم تفاوتی ندارند.

**جدول ۳.** نتایج مقایسه رتبه‌بندی مقاله‌های مرتبط پیشنهادی پایگاه وب‌آوساینس و گوگل‌اسکالر براساس سنجه NDCG

| | میانگین رتبه‌ها | Zآماره | سطح معناداری |
|---|---|---|---|
| NDCG وب‌آوساینس | ۵۸/۸۱ | -۰/۶۹۹ | ۰/۴۸۵ |
| NDCG گوگل‌اسکالر | ۵۸/۱۴ | | |

## نتیجه‌گیری

در هر دو پایگاه وب‌آوساینس و گوگل‌اسکالر میان رتبه‌بندی کاربران و رتبه‌بندی سامانه همبستگی ضعیف بود. تقدم و تأخر مقالات مرتبط پیشنهادی، از نظر کاربران با رتبه‌ای که پایگاه به آنها اختصاص داده بود، نزدیک نبود. این احتمال وجود دارد که نخستین مقاله موجود در این فهرست، به‌رغم تشخیص الگوریتم‌های این دو پایگاه به‌عنوان مرتبط‌ترین مقاله و در صدر فهرست نتایج و در رتبه اول بازیابی‌شده شود، اما کاربر آن را به‌عنوان چندمین مقاله رتبه‌بندی کند؛ و برعکس، مقاله قرارگرفته در انتهای فهرست مقاله‌های مرتبط پیشنهادی، به‌عنوان مرتبط‌ترین مقاله و با رتبه یک توسط کاربر رتبه‌بندی شود. همچنین، نتیجه تحلیل رتبه‌بندی دو پایگاه با استفاده از شاخص NDCG نشان داد این دو پایگاه به‌لحاظ عملکرد رتبه‌بندی مدارک مرتبط پیشنهادی همبستگی ضعیفی دارند. نکته مهم و درخور توجه این است که این پژوهش روی ۱۰ نتیجه نخست فهرست نتایج انجام شد و چنانچه پژوهش روی نتایج بعدی بازیابی (بعد از ۱۰ نتیجه نخست) انجام شود، این احتمال وجود دارد که ضریب/های همبستگی ضعیف‌تر یا حتی بی‌معنا باشد.

پژوهشی که الگوریتم رتبه‌بندی مقاله‌های مرتبطِ پیشنهادی موتور جستجوی گوگل‌اسکالر را بررسی کرده باشد یافت نشد جز درباره پایگاه پاب‌مد (Lingeman & Yu, 2016) که نشان می‌داد میان رتبه‌ای که سامانه به مقاله‌های مرتبط پیشنهادی



www.SID.ir

Archive of SID

اختصاص داده بود و رتبه‌ای که کاربران به همان مقالات اختصاص داده بودند تفاوت معنادار وجود دارد.

گوگل‌اسکالر از معدود موتورهای جستجو است که ترکیبی از چند رویکرد (متنی و استنادی) را در الگوریتم خود استفاده می‌کند (Beel & Gipp, 2009؛ آتشکار، علیپورحافظی، و نوروزی، ۱۳۹۲). در این شیوه، تعداد استنادهایی که یک مقاله دریافت می‌کند در الگوریتم رتبه‌بندی گوگل‌اسکالر مهم‌ترین عامل رتبه‌بندی است (Martín-Martín, Orduña-Malea, Ayllón, & López-Cózar, 2014)؛ ولی تنها عامل نیست. بود یا نبود اصطلاح پرسش در عنوان مقاله، فراوانی اصطلاح پرسش در متن کامل مقاله، عمر مقاله[1] و مطابقت اصطلاح پرسش با نام نویسنده و نام مجله تأثیرگذارند. الگوریتم رتبه‌بندی با الگوریتم استفاده برای رتبه‌بندی متفاوت است. کاربران، مقاله‌های مرتبط پیشنهادی در موتور جستجوی گوگل‌اسکالر را اثربخش دانسته‌اند (سعدین، عباس‌پور، و ستوده، زودآیند)؛ اما قرارگرفتن مقاله‌های کاملاً مرتبط در صدر نتایج جستجو، برای کاربران مهم است. همبستگی ضعیف میان رتبه‌بندی کاربران و سامانه نشان می‌دهد الگوریتم‌ها و شاخص‌های گوگل‌اسکالر برای رتبه‌بندی مقاله‌های مرتبط در برآوردن نیازهای کاربران چندان موفق نیست.

الگوریتم‌های رتبه‌بندی پایگاه وب‌آوساینس از سنجه اشتراک در مآخذ برای پیشنهاد مدارک مرتبط به کاربران استفاده می‌کند (Char & Ajiferuke, 2013). سنجه اشتراک در مآخذ استنادمحور است و برمبنای آن هرچه اشتراک مآخذ دو مدرک بیشتر باشد شبیه‌تر تلقی می‌شوند و در جایگاه بالاتری قرار می‌گیرند. در این پژوهش، ۱۰ مقاله نخست پیشنهادشده که انتظار می‌رود بهترین و مرتبط‌ترین مقالات پیشنهادشده پایگاه باشند بررسی شد. با وجود آنکه از نظر کاربران، پایگاه وب‌آوساینس بین ۱۰ نتیجه نخست، تعداد قابل‌قبولی مدارک مرتبط و تاحدی مرتبط را برای کاربران بازیابی می‌کند (سعدین، عباس‌پور، و ستوده، زودآیند)، جایگاه و ترتیب قرارگرفتن این ۱۰ مقاله هم مهم و لازم است پایگاه در بین ۱۰ نتیجه بازیابی‌شده، تعداد قابل‌قبول مدارک مرتبط را بازیابی کند. در کنار آن بایستی این مدارک به‌ترتیب اهمیت در بالاترین جایگاه قرار گیرند و الگوریتم‌های به‌کارگرفته‌شده در پایگاه آنقدر قوی باشد که بتواند چنین امکانی را فراهم کند. این پژوهش و میزان ضریب همبستگی به‌دست‌آمده نشان داد الگوریتم‌ها و شاخص‌های پایگاه وب‌آوساینس برای رتبه‌بندی نتوانسته است رضایت کامل کاربران را تأمین کند.

نتایج پژوهش از یک‌سو و اینکه دو پایگاه بررسی‌شده بیش از هر عاملی به شاخص‌های استنادی برای رتبه‌بندی مقاله‌های پیشنهادی اتکا می‌کنند، لازم می‌آورد

1. Article's age


www.SID.ir

Archive of SID



در کارایی رویکرد استنادمحور، یا دست‌کم وزن آن در الگوریتم‌ها تجدیدنظر شود. البته برای تأیید این نظر و تعیین سهم هریک از شاخص‌های استنادی و متنی در الگوریتم‌ها به پژوهش‌های بیشتر و کنترل عوامل مرتبط و دخیل نیاز است.

## مآخذ


آتشکار، مرضیه؛ علیپورحافظی، مهدی؛ و نوروزی، یعقوب (۱۳۹۲). شناسایی میزان آشنایی دانشجویان تحصیلات تکمیلی با پایگاه‌های گوگل‌اسکالر. *نظام‌ها و خدمات اطلاعاتی*، ۹ (۱)، ۶۱-۷۸.

دلاور، علی (۱۳۹۶). *روش تحقیق در روانشناسی و علوم تربیتی* (ویرایش ۴). تهران: ویرایش.

ره‌گشای، مرتضی (۱۳۹۰). *مطالعه ابرموتورهای جستجو در پاسخ‌گویی به سؤالات کاربران کتابداری و اطلاع‌رسانی و ارایه الگوی پیشنهادی جهت بهبود رتبه‌بندی نتایج جستجو*. پایان‌نامه کارشناسی ارشد، دانشگاه پیام نور، مشهد.

ریاحی‌نیا، نصرت؛ رحیمی، فروغ، لطیفی، معصومه؛ و بخشیان، لیلی‌الله. (۱۳۹۴). بررسی میزان انطباق ربط سیستمی و ربط کاربرمدارانه در پایگاه‌های اطلاعاتی Google Scholar - ISC - SID. *تعامل انسان و اطلاعات*، ۲ (۱)، ۱-۱۱.

سعدین، صبا؛ عباس‌پور، جواد؛ و ستوده، هاجر (زودآیند). مقایسه اثربخشی سامانه‌های پیشنهاددهنده مقاله‌های مرتبط در پایگاه‌های وب‌آوساینس و گوگل‌اسکالر. *تحقیقات کتابداری و اطلاع‌رسانی دانشگاهی*.

فرهودی، فائزه؛ حریری، نجلا (۱۳۹۲). تأثیر ویژگی‌های شخصیتی کاربران بر قضاوت ربط. *پردازش و مدیریت اطلاعات*، ۲۹ (۲)، ۳۱۷-۳۳۱.

Al-Maskari, A., Sanderson, M., & Clough, P. (2007). The relationship between IR effectiveness measures and user satisfaction. In *Proceedings of the* 30*th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, *July* 23-27, (pp. 773-774). Retrived October 30, 2019, from http://www.marksanderson.org/publications/my_papers/SIGIR2007-b.pdf.

Bar-Ilan, J., Keenoy, K., Levene, M., & Yaari, E. (2009). Presentation bias is significant in determining user preference for search results—A user study. *Journal of the American Society for Information Science and Technology*, 60 (1), 135-149.

Bar-Ilan, J., Keenoy, K., Yaari, E., & Levene, M. (2007). User rankings of




www.SID.ir

Archive of SID

search engine results. *Journal of the American Society for Information Science and Technology*, 58 (9), 1254-1266.

Bar-Ilan, J., Levene, M., & Mat-Hassan, M. (2006). Methods for evaluating dynamic changes in search engine rankings: a case study. *Journal of Documentation*, 62 (6), 708-729.

Beel, J., & Gipp, B. (2009). Google scholar's ranking algorithm: The impact of articles' age (an empirical study). In S. Latifi (Ed.), *Proceedings of the* 6*th International Conference on Information Technology*: *New Generations*, *April* 27-29, (pp. 160-164). IEEE. Retrieved October 30, 2019, from https://ieeexplore.ieee.org/document/5070610

Beg, M. S. (2005). A subjective measure of web search quality. *Information Sciences*, 169 (3-4), 365-381.

Char, D. C., & Ajiferuke, I. (2013, October). Comparison of the effectiveness of related functions in Web of Science and Scopus. In *Proceedings of the Annual Conference of CAIS/Actes du Congrès Annuel de l'ACSI*. Retrieved October 30, 2019, from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=12BDD2E4D8D78A77 7A4BCDD5E8 FD38B1?doi=10.1.1.181.382&rep=rep1&type=pdf

Drori, O. (2002). Algorithm for documents ranking: Idea and simulation results. In *Proceedings of the* 14*th international conference on Software Engineering and Knowledge Engineering*, *July* 15-19, (pp. 99-102). New York, NY: ACM.

Eto, M. (2013). Evaluations of context-based co-citation searching. *Scientometrics*, 94 (2), 651-673.

Hariri, N. (2011). Relevance ranking on Google: Are top ranked results really considered more relevant by the users? *Online Information Review*, 35 (4), 598-610.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing & Management*, 36 (2), 207-227.



www.SID.ir

Archive of SID



Joachims, T., Granka, L., Pan, B., Hembrooke, H., & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, August 15-19, (pp. 154-161). New York, NY: ACM.

Kekäläinen, J. (2005). Binary and graded relevance in IR evaluations—comparison of the effects on ranking of IR systems. *Information processing & management*, 41 (5), 1019-1033.

Kinley, K., Tjondronegoro, D., Partridge, H., & Edwards, S. (2014). Modeling users' web search behavior and their cognitive styles. *Journal of the Association for Information Science and Technology*, 65 (6), 1107-1123.

Lewandowski, D. (2008). The retrieval effectiveness of web search engines: considering results descriptions. *Journal of Documentation*, 64 (6), 915-937.

Lingeman, J. M., & Yu, H. (2016). Learning to Rank Scientific Documents from the Crowd. *arXiv preprint arXiv*:1611.01400. Retrieved October 30, 2019, from https://arxiv.org/pdf/1611.01400.pdf

Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M., & López-Cózar, E. D. (2014). Does Google Scholar contain all highly cited documents (1950-2013)? Retrieved October 30, 2019, from https://arxiv.org/ftp/arxiv/papers/1410/1410.8464.pdf

Nowicki, S. (2003). Student vs. search engine: Undergraduates rank results for relevance. *Portal*: *Libraries and the Academy*, 3 (3), 503-515.

Patil, S., Alpert, S. R., Karat, J., & Wolf, C. (2005,). "THAT's what i was looking for": Comparing user-rated relevance with search engine rankings. In *IFIP Conference on Human-Computer Interaction*, *September* 12-16, (pp. 117-129). Berlin, Heidelberg: Springer.

Sakai, T. (2007, May). On Penalising Late Arrival of Relevant Documents in Information Retrieval Evaluation with Graded Relevance. In *The First International Workshop on Evaluating Information Access* (*EVIA*), *May* 15, (pp. 32-43). Retrieved October 30, 2019, from http://research.nii.ac.jp/



www.SID.ir

Archive of SID

ntcir/ntcir-ws6/OnlineProceedings/EVIA_Preprint_Papers/1.pdf

Sanderson, M., Paramita, M. L., Clough, P., & Kanoulas, E. (2010, July). Do user preferences and evaluation measures line up? In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, (pp. 555-562). Retrieved October 30, 2019, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.190.9251&rep=rep1&type=pdf

Su, L. T. (2003). A comprehensive and systematic model of user evaluation of Web search engines: II. An evaluation by undergraduates. *Journal of the American Society for Information Science and Technology*, 54 (13), 1193-1223.

Teevan, J., Dumais, S. T., & Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, *August* 15-19, (pp. 449-456). New York, NY: ACM.

Vallez, M., & Pedraza-Jimenez, R. (2007). Natural language processing in textual information retrieval and related topics. *Hypertext.net*, 5. Retrieved October 30, 2019, from https://www.upf.edu/hipertextnet/en/numero-5/pln.html

Wang, Y., Wang, L., Li, Y., He, D., & Liu, T. Y. (2013, June). A theoretical analysis of NDCG type ranking measures. In *Conference on Learning Theory*, *June* 12-14, (pp. 25-54). Retrieved October 30, 2019, from http://proceedings.mlr.press/v30/Wang13.pdf

Yoon, S. H., Kim, S. W., Kim, J. S., & Hwang, W. S. (2011). On computing text-based similarity in scientific literature. In *Proceedings of the 20th International Conference Companion on World Wide Web*, March 28 - April 1, (pp. 169-170). New York, NY: ACM.

## استناد به این مقاله:

سعدین خرم، صبا؛ عباسپور، جواد (۱۳۹۸). سنجش رتبه‌بندی سامانه‌های پیشنهاددهندهٔ مقاله در تقابل با رتبه‌بندی کاربران. *مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۳۰ (۳)، ۴۶-۵۷.



www.SID.ir