

Citation Analysis of Accounting and Auditing Research in Iran: A Text Mining Approach

Hamideh Asnaashari

Assistant Professor of Accounting, Management and Accounting Faculty, Shahid Beheshti University, Tehran, Iran

Mohammad Sadra Moradnia*

MSc. Student of accounting, Management and Accounting Faculty, Shahid Beheshti University, Tehran, Iran.

Sina Asnaashari

Ph.D. student of accounting, Management and Accounting Faculty, Shahid Beheshti University, Tehran, Iran.

Abstract

Given that the number of citations a research work gives to other studies is considered a measure of its scientific credibility, researchers might focus more on the quantity of citations rather than their thematic relevance to their own work. Therefore, this study aims to examine the thematic similarity between the citations and the core topic of the citing research. To achieve this objective, Natural Language Processing (NLP) was employed as one of the most significant approaches for working with textual data in machine learning. After extracting citations data from Science Ministry journals' web page since 1393 until 1402 using web crawling, thematic relevance of citations and the core theme of each study was calculated using allmpnet model. The results indicate that the thematic cohesion between citations and the core research topic in Iranian journals is at a proper level. Furthermore, the citations by Iranian researchers are primarily to sources published in first and second-quartile journals (Q1) and (Q2), indicating their credibility. Additionally, the most highly-cited sources in Iranian accounting and auditing research were identified. The novel methodology and findings of this study can significantly contribute to evaluating citation practices. It can assist journals in more accurately assessing citations, helping them move beyond a sole focus on citation count and instead utilize the degree of citation relevance as an indicator for evaluating citation quality.

Keywords: citation analysis, natural language processing, machine learning, text mining, Bidirectional Encoder Representations from Transformers (BERT)

* Hamideh Asnaashari:h_asnaashari@sbu.ac.ir

How to Cite: xxxxxxxx

1. Introduction

The number of citations a paper receives has become one of the most visible proxies for scientific credibility in modern academia. Journals, funding agencies, promotion committees, and even national ranking systems routinely treat citation counts as evidence of impact (Dechow et al., 2020; van Raan et al., 2003; Woolgar, 1991). Yet this very pressure creates a perverse incentive: researchers may inflate reference lists with prestigious but topically distant sources simply to boost perceived legitimacy (Dai et al., 2021; Iqbal et al., 2021).

In this study we define citation quality through a new lens: thematic cohesion, the degree of semantic alignment between the core topic of a paper and the titles of every work it cites. Although erroneous and irrelevant citations have been documented even in top-tier international journals (Evans et al., 1990), no large-scale, automated assessment of thematic cohesion has ever been conducted in Iranian accounting and auditing literature.

Using web scraping, Scopus quartile mapping, and state-of-the-art Sentence-BERT embeddings, we address three interrelated questions:

- To what extent do Iranian accounting researchers cite journals of established global prestige?
- How strong is the thematic cohesion between Iranian papers and the works they cite?

By moving beyond raw counts and making relevance visible, measurable, and enforceable, this study offers editors a practical tool to reward substance over symbolism, and provides Iranian accounting scholarship with a roadmap to even greater international influence.

2. Methodology

This quantitative study adopts an applied research design that combines content analysis with large-scale text mining. The population comprises all references appearing in articles published in Iran's Ministry of Science-approved accounting and auditing journals between 2014 and 2023. Of the 20 recognized journals, 14 provided full-text HTML archives suitable for automated retrieval, forming the final sampling frame.

Data were collected using custom Python web crawlers built with Requests and BeautifulSoup libraries. The scripts systematically traversed journal archives, extracted article-level metadata, and downloaded full texts, yielding 3,436 published papers. After excluding 470 articles whose PDFs could not be parsed, 2,966 full texts entered the pipeline. From these, 132,543 raw references were harvested. Four exclusion filters were then applied including Persian-language references, books, theses, conference papers, and working papers, legal documents and standards, and non-journal items without DOI or

Scopus entry. This rigorous cleaning retained 65,146 foreign journal citations from 2,938 articles.

Each cited journal was matched to its ISSN in Scopus Source List (2023 release). Quartiles were assigned dynamically by year and subject category, producing the final distribution: Q1 (77.5%), Q2 (15.2%), Q3 (5.4%), Q4 + unranked (2.0%).

Semantic similarity between every citing-title / cited-title pair was computed using multilingual Sentence-BERT (sentence-transformers/paraphrase-multilingual-mpnet-base-v2), fine-tuned on more than 100 languages including Persian and English. Cosine similarity scores were scaled to 0–100 percent. Inter-coder reliability was established on a random 500-pair subsample manually labelled by two accounting PhD candidates. Embedding stability was confirmed by repeating the analysis with three different seeds; maximum variation in journal-level means was 0.12 percentage points.

By combining web-scale data acquisition with state-of-the-art NLP, this methodology delivers the first objective, replicable measure of citation relevance ever applied to an entire national accounting literature.

3. Results

The analysis of 65,146 international journal citations reveals a clear portrait of Iranian accounting scholarship: globally oriented, prestige-conscious, and already practising respectable thematic discipline. Iranian papers' citations suggest that 77 percent of the citations are cited to Q1 journals. No ministry-approved journal falls below 67% Q1+Q2 combined. The leader, "Professional Auditing Research", directs 83.1% of its citations to Q1 outlets; even the most locally oriented title, "Governmental Accounting", still achieves 53.7% Q1.

Across all 65,146 citation pairs, the mean thematic cohesion is 45.4%. The distribution is remarkably symmetrical and bell-shaped, peaking at 45–50%. Only 9% of citations reach the high-to-excellent zone above 55%, while 10% sink into weak territory below 35%. In other words, Iranian accounting researchers rarely cite completely irrelevant work, but they also rarely achieve perfect topical harmony.

Every single year stays within the 44.7–47.6% corridor. The absence of a strong upward or downward trend suggests that citation culture has matured but not yet entered a phase of deliberate relevance optimisation.

Taken together, the results paint Iranian accounting scholarship as internationally ambitious and methodologically disciplined. The community already cites the right journals; the next frontier is citing the right ideas inside them.

4. Discussion

The fact that average thematic cohesion plateaus at 45.4 %—solidly "good" yet never crossing the 50 %, reveals a subtle but systematic gap. Ten percent

of all citations drift into the weak zone (<35 %). Another 81 % hover in the moderate-to-good band, useful but rarely transformative. Only 9 % achieve the high-to-excellent alignment that signals a genuine intellectual dialogue. This gap is not unique to Iran. Global studies repeatedly warn that citation inflation, Matthew Effect, and strategic referencing are pandemic (Biglu, 2008; Bornmann & Daniel, 2008; Dai et al., 2021). What makes the Iranian case instructive is the stability of the 45 % ceiling across ten years and fourteen journals.

The benefit of thematic cohesion is that it is cheap, instantaneous, and impossible to game. A single Sentence-BERT inference on a standard laptop takes 0.03 seconds per citation pair. An entire manuscript can be scored before the reviewer finishes the first page. Journals that adopt a modest gate, would immediately filter out decorative references while preserving author freedom. Early adopters elsewhere have reported reference list shrinkage of 12–18 % with no loss of scientific depth (Aljuaid et al., 2021; Anglin et al., 2021).

5. Conclusion

Iranian accounting and auditing scholarship has reached a pivotal moment. For a decade, the community has looked outward with confidence: 77.5 % of all citations now flow to Q1 journals. Yet global engagement must now evolve into global conversation. The 45.4 % average thematic cohesion achieved across 65,146 citations tells editors that most references already serve the argument, but one in ten does not, and only one in eleven truly transforms it.

When Iranian journals lead the shift toward the new article evaluation measure provided in this paper, first, domestic papers will become sharper, more readable, and more persuasive. Second, international reviewers will notice the difference and reward it with faster acceptances and higher citations. The future of Iranian accounting research is therefore not about citing more; it is about citing better.

تحلیل استنادی پژوهش‌های حسابداری و حسابرسی در ایران: رویکرد متن کاوی

استادیار، گروه حسابداری، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران.

حمیده اثنی

عشری *

دانشجوی کارشناسی ارشد، رشته حسابداری، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران.

محمد صدرا مرادنیا



دانشجوی دکتری، رشته حسابداری، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران.

سینا اثنی عشری 

چکیده

با توجه به اینکه تعداد استناداتی که یک پژوهش به سایر پژوهش‌ها می‌دهد، به عنوان یکی از معیارهای اعتبار علمی آن تلقی می‌شود، ممکن است پژوهشگران به تعداد استنادات، بیش از ارتباط موضوعی آن به پژوهش خود توجه نمایند؛ از این رو، این پژوهش قصد دارد تا به تحلیل شباهت موضوعی ارجاعات پژوهش‌های داخلی با موضوع محوری پژوهش پردازد تا میزان مربوط بودن استنادات مورد استفاده به موضوع پژوهش در پژوهش‌های داخلی مشخص گردد. به منظور دستیابی به این هدف از پردازش زبان طبیعی به عنوان یکی از مهم‌ترین رهیافت‌های کار با داده متنی در یادگیری ماشین استفاده شده است. پس از استخراج اطلاعات مقالات نشریات وزارت علوم طی سال‌های ۱۳۹۳ لغایت ۱۴۰۲ توسط روش خزش وب، شباهت موضوعی میان استنادات خارجی پژوهش‌های داخلی با موضوع محوری پژوهش با استفاده از

* حمیده اثنی عشری: h_asnaashari@sbu.ac.ir

مدل آل ام پی نت بدست آمد. نتایج حاکی از آن بود که چسبندگی موضوعی میان استنادات و موضوع محوری پژوهش به طور کلی در نشریات ایرانی در سطح خوب قرار دارد. همچنین ارجاعات پژوهشگران ایرانی عمدتاً به منابع منتشر شده در نشریات چارک‌های نخست و دوم بوده لذا معتبر هستند. افزون بر آن، پر استنادترین منابع در پژوهش‌های حوزه حسابداری و حسابرسی ایران شناسایی شدند. روش‌شناسی نوین و نتایج این پژوهش می‌تواند کمک شایانی به ارزیابی شیوه ارجاع‌دهی کند و نشریات را در ارزیابی دقیق‌تر استنادات یاری رساند تا از تمرکز صرف بر تعداد استناد بکاهند و میزان ارتباط استنادات را با موضوع محوری پژوهش را نیز به عنوان شاخصی جهت ارزیابی شیوه ارجاع‌دهی استفاده نمایند.

کلیدواژه‌ها: تحلیل استناد، یادگیری ماشین، پردازش زبان طبیعی، متن‌کاوی، نمایش رمزگذار دوسویه از تبدیل‌کننده‌ها (BERT)

راهنمای انتشار

مقدمه

پژوهش‌هایی که به بررسی رفتارهای ارجاع‌دهی میان دانشمندان پرداخته‌اند نشان می‌دهند که ارجاع‌دهی‌ها الزاماً دلایل علمی ندارند و عوامل گوناگون می‌توانند بر رفتار ارجاع‌دهی موثر باشد (Bornmann & Daniel, 2008). به این ترتیب رفتارهای غیرعلمی ممکن است به ارجاعات کم کیفیت‌تر و کم ارتباط به پژوهش منتج شود؛ به عبارتی دیگر، پژوهشگران را به سوی ارجاع‌دهی به پژوهش‌های زیاد و در عین حال کم ارتباط به پژوهش سوق دهد.

ارتباط موضوعی میان پژوهش و ارجاعات آن را در این پژوهش چسبندگی موضوعی می‌نامیم؛ به بیانی دیگر، چسبندگی موضوعی نشان می‌دهد تا چه میزان موضوع استنادات صورت گرفته در یک پژوهش با موضوع محوری آن پژوهش مرتبط است. هرچه چسبندگی پایین‌تر باشد، ارجاعات مرتبط‌تر با موضوع پژوهش خواهد بود؛ و بالعکس، چسبندگی پایین نیز بیانگر ارتباط کمتر ارجاعات با موضوع پژوهش است. با توجه به اینکه تعداد استنادات هر پژوهش، یک معیار خام برای اندازه‌گیری اعتبار و علمی بودن آن پژوهش به شمار می‌رود (Dai et al., 2021; van Raan et al., 2003; Woolgar, 1991)، پژوهشگران ممکن است نه به دلیل ارتباط مناسب منبع، بلکه به دلیل اعتبار منبع، به آن ارجاع دهند تا به پژوهش خود اعتبار بخشند (Iqbal et al., 2021). کاهش چسبندگی موضوعی ناشی از ارجاع‌دهی به تعداد بالا به پژوهش‌های بی‌ربط به موضوع پژوهش، باعث کاهش چسبندگی موضوعی و همچنین کیفیت استنادات می‌شود. به این ترتیب چسبندگی موضوعی استنادات، می‌تواند به عنوان یک معیار برای اندازه‌گیری کیفیت استنادات و در پی آن، اعتبار کلی پژوهش مورد استفاده قرار گیرد.

چسبندگی موضوعی میان ارجاعات و موضوع محوری پژوهش از آن جهت اهمیت دارد که انتقال دانش میان علوم، مادامی که ارتباط مناسبی میان استنادات و موضوع پژوهش وجود ندارد، به درستی صورت نمی‌گیرد. با این وجود، حتی در نشریات معتبر خارجی در علوم مختلف، تحلیل استنادات نشان داده است، ارجاعات اشتباه امری رایج است (Evans et al., 1990). بنابراین اندازه‌گیری چسبندگی موضوعی می‌تواند در پژوهش‌ها می‌تواند به کاهش ارجاعات اشتباه، با شناسایی استنادات غیر مرتبط به موضوع محوری پژوهش، بسیار کاردردی باشد.

با توجه به اینکه در ادبیات پژوهشی داخلی به نگرانی‌های موجود در خصوص چسبندگی موضوعی در ارجاع‌دهی پاسخ داده نشده است، این پژوهش به بررسی و تجزیه و تحلیل ارجاعات خارجی پژوهش‌های داخلی صورت گرفته در حسابداری و حسابرسی پرداخته است تا از این طریق، روندها و الگوهای استناد به منابع خارجی استخراج، و نشریات پر استناد و در نهایت چسبندگی موضوعی مورد تجزیه و تحلیل قرار گیرند.

این پژوهش از متن کاوی^۱ و پردازش زبان طبیعی^۲ به عنوان دو ریکرد نوین در فناوری اطلاعات برای تحلیل داده‌های متنی و دستیابی به اهداف پژوهش استفاده می‌کند. این روش‌ها افزون بر افزایش دقت و سرعت پردازش داده‌های متنی، به شناسایی الگوها و روابط پنهان در منابع علمی کمک می‌کند که تا پیش از این ابزارها، ممکن نبودند (Shi, 2024). بنابراین ارائه راهکاری جهت سنجش چسبندگی موضوعی استنادات، به جای تکیه صرف بر تعداد استنادات، دانش‌افزایی اصلی این پژوهش است. نتایج این پژوهش به بهبود کیفیت پژوهش‌های داخلی و رقابت مقالات علمی، بهبود جهت‌دهی به پژوهش‌های نوین و سیاست‌گذاران پژوهشی، مجلات و سردبیران آن‌ها و خود پژوهشگران به منظور شناسایی ارجاعات مناسب برای توسعه علم حسابداری کمک می‌کند.

در ادامه پژوهش، مبانی نظری به تحلیل استنادی، چسبندگی موضوعی و رویکردهای تحلیل استنادی پرداخته می‌شود. سپس روش‌شناسی پژوهش تشریح، یافته‌ها ارائه و در نهایت یافته‌ها مورد بحث و نتیجه‌گیری قرار می‌گیرند.

مبانی نظری و پیشینه پژوهش

تحلیل استنادی

اهمیت استنادها، فراتر از تأیید پژوهش‌های گذشته است و نقش مهمی در ایجاد پیوندهای فکری میان رشته‌های مختلف و حوزه‌های جغرافیایی گوناگون (Silvello, 2018) ایفا می‌کنند و نشان‌دهنده تاثیر یافته‌ها در عمل (García Osmá et al., 2023)، ردگیری سیر تکاملی دانش (Skolarus et al., 2017)، بهبود انتشار دانش و دسترسی به منابع (Andrew et al., 2024) هستند. تحلیل استنادی روشی علم‌سنجی است که اثرگذاری

Text mining^۱
Natural language processing (NLP)^۲

پژوهش‌های علمی را از طریق تحلیل استنادات می‌سنجد. همچنین معیار است برای سنجش کیفیت و اعتبار آثار علمی (Dechow et al., 2020; Meyer et al., 2018). هر چند نگرانی‌هایی نیز نسبت به تمرکز بیش از حد بر استنادها به عنوان معیار کیفیت مطرح است؛ چراکه خود استنادی و ارجاعات میان‌رشته‌ای هر دو می‌توانند بر تعداد استنادها تاثیرگذار باشند (Kim et al., 2015). تحلیل استنادی همچنین می‌تواند در شناسایی تاثیر نشریات بین‌المللی بر نشریات محلی بکار گرفته شود. افزون بر آن به پیش‌بینی پاسخگویی جوامع علمی به موضوعات نوظهور از طریق بررسی تاخیر زمانی انتشار این موضوعات در پژوهش‌های محلی نسبت به بین‌المللی کمک می‌کند (Cho et al., 2022). از چالش‌های این تحلیل، بکار نبردن شیوه‌های ارجاع‌دهی استاندارد و قالب‌های یکسان اشاره کرد (Iqbal et al., 2021). تحلیل استنادی در حسابداری و حسابرسی به خصوص در موضوعاتی مانند کیفیت گزارشگری مالی، استانداردهای حسابداری، حاکمیت شرکتی و حسابرسی قانونی (رعایت) که ارجاعات بالایی دریافت می‌کنند (Khelif & Chalmers, 2015)، با استفاده از روش‌هایی همچون کتاب‌سنجی (مانند Behrend & Eulerich, 2019) مورد قرار گرفته است. همچنین با تحلیل استنادها، نشریات و موسسات اثرگذار نیز شناسایی می‌شوند که منتج به شناسایی معیاری جهت ارزیابی این نشریات و موسسات می‌شود (de Villiers & Hsiao, 2018).

عواملی که بر تعداد استنادات موثر هستند همواره حوزه مورد علاقه پژوهشگران بوده است. به عنوان مثال پژوهش‌هایی که مبنای نظری مستحکم‌تر، روش‌شناسی دقیق‌تری دارند و موضوعات نوظهورتری را بررسی می‌کنند، به علت پایه‌گذاری مستحکم‌تر برای پژوهش‌های آتی، استناد بیشتری جذب می‌کنند (Dechow et al., 2020; Jones & Alam, 2019). نشریاتی که در رتبه بندی، در چارک نخست قرار می‌گیرند، و مجلاتی که دسترسی آزاد برای پژوهشگران فراهم می‌کنند، به دلیل بیشتر دیده شدن، استناد بیشتری جذب می‌کنند (Onodera & Yoshikane, 2015; Sangster et al., 2015). شهرت نویسنده و پژوهشکده‌ها به دلیل نفوذ بیشتر در جامعه علمی، عامل دیگر در جذب استنادات است (Myers et al., 2016; Negash et al., 2019). همکاری نویسندگان به صورت بین‌المللی، بر استنادات دریافتی تاثیر مثبت دارد (Endenich & Trapp, 2016).

موضوعات پژوهشی نیز بر استنادها موثر هستند. به طور کلی برخی حیطه‌های پژوهشی مانند حسابداری و حسابرسی به دلیل داشتن جامعه پژوهشی کوچکتر و چرخه‌های انتشار طولانی‌تر نسبت به رشته‌هایی مانند مهندسی و پزشکی، نرخ استناد کمتری دارند (Myers et al., 2016). حوزه جغرافیایی نیز به دلیل شبکه‌های پژوهشی ایجاد شده توسط مراکز پژوهشی، دیگر عامل موثر بر تعداد استنادات است (de Villiers & Hsiao, 2018).

در حسابداری و حسابرسی، از بابت موضوع، حوزه‌های گزارشگری مالی، حاکمیت شرکتی، کیفیت حسابرسی و مالیات و از بابت زمان، نشریات قدیمی‌تر و اثرگذار در حسابداری، استناد بیشتری کسب می‌کنند (Meyer et al., 2018; Myers et al., 2016). Mashayekhi et al. (2024) نیز در بررسی استناد پژوهش‌های حسابرسی، دریافتند که موضوعات مورد استناد عموماً مربوط به کیفیت حسابرسی، استقلال حسابرس و کشف تقلب مالی است.

چسبندگی موضوعی پژوهش

چالش‌های بسیاری درباره ارجاع‌دهی و استنادها وجود دارند. به عنوان مثال، خود استنادی، دستکاری استنادها، توزیع نابرابر استنادها میان رشته‌های گوناگون و نگرانی درباره قابلیت اتکای ارجاعات از چالش‌های انتشار دانش از طریق استنادات است (Dechow et al., 2020). بنابراین تکیه صرف بر تعداد استنادات معیار ارزیابی کیفیت مناسبی نیست (Aljuaid et al., 2021)؛ چراکه باعث می‌شود تا نویسندگان، بی‌توجه به ارتباط موضوعی پژوهش‌های مورد استناد، تمایل بیشتری به افزایش تعداد استنادات خود داشته باشند. این تمایل بیشتر ناشی از این دیدگاه است که تعداد ارجاعات بالا، به عنوان معیاری از تلاش‌های پژوهشگر، احتمال انتشار و دریافت ارجاعات را افزایش می‌دهد (Dai et al., 2021). Iqbal et al. (2021) همچنین بیان می‌کنند که پژوهشگران گاهی ممکن است با هدف افزایش اعتبار پژوهش خود، به منابع معتبر ارجاع دهند. این موضوع ممکن است نویسندگان را به ارجاع به منابع غیر مرتبط نیز سوق دهد.

تعداد ارجاعات در علوم مختلف از اهمیت بالایی برخوردار است؛ به طور مثال (Biglu, 2008) تعداد ارجاعات را طی سال‌های ۱۹۷۰ تا ۲۰۰۵ مورد بررسی قرار داده است و افزایش

قابل توجه تعداد ارجاعات را با اثر متیو^۱ توجیه کرد. اثر متیو در این حوزه پژوهشی به این معنی است که نویسندگان شناخته شده‌تر، راحت‌تر توسط نشریات مورد پذیرش قرار می‌گیرند؛ به همین دلیل، ارجاعات به ایشان نیز به افزایش احتمال پذیرش کمک می‌کند. بنابراین اثر متیو می‌تواند نویسندگان را به ارجاع‌دهی به نویسندگان شناخته شده‌تر، حتی در صورت کم ارتباط بودن به موضوع محوری پژوهش، سوق دهد.

چسبندگی موضوعی در این پژوهش به معنی شباهت اسنادها به موضوع محوری پژوهش است. تحلیل چسبندگی موضوعی، به شناسایی خود استنادی و اسنادهای نامرتب توسط پژوهشگران کمک کند تا کاستی ناشی از تعداد اسنادها به عنوان معیار اعتبار پژوهش به میزان مناسبی مرتفع گردد (Sangster et al., 2015). از این رو پردازش زبان طبیعی و یادگیری ماشین می‌تواند بر ناکارآمدی رویکردهای سنتی کتاب‌سنجی که عمدتاً به شمار اسنادها و ضریب تأثیر به عنوان معیاری برای اعتبار اتکا می‌کنند چیره شوند و هم‌راستایی اسنادها و موضوع پژوهش را نیز در نظر بگیرند (Anglin et al., 2021).

رویکردهای روش‌شناختی در تحلیل استنادی

در تحلیل استنادی به دلیل نیاز به بررسی تعداد بالایی از داده‌های متنی، در سال‌های اخیر استفاده از پردازش زبان طبیعی به عنوان شاخه‌ای از یادگیری ماشین، در تحلیل استنادی جای خود را در پژوهش‌ها باز کرده است (Iqbal et al., 2021). تکنیک‌های مختلف داده‌کاوی برای شناسایی و دسته‌بندی اسنادها، تحلیل احساسات و ارزیابی حوزه‌های جستجوی مرتبط مانند سیستم‌های پیشنهادی مبتنی بر استناد و خلاصه‌سازی مبتنی بر استناد (Iqbal et al., 2021)، نمونه‌هایی از کاربردهای الگوریتم‌های مختلف برای پردازش کلان‌داده‌های متنی در تحلیلی استنادی هستند که پیش از پیشرفت‌های اخیر در فناوری دور از ذهن به نظر می‌رسیدند.

برای استفاده از الگوریتم‌های گوناگون در تحلیل استنادی، ابتدا نیاز به گردآوری داده‌های مربوط به ارجاعات است. یکی از این راه‌ها، استخراج این داده‌ها از صفحات وب نشریات با استفاده از وب اسکرپینگ^۲ است. وب اسکرپینگ، با استفاده از زبان‌های برنامه‌نویسی،

استخراج داده‌ها را از صفحات وب و مخازن داده‌های دیجیتال به صورت خودکار و ساماندهی شده امکان‌پذیر می‌کند (Boegershausen et al., 2022; Li et al., 2021). مراحل وب اسکرپینگ شامل شناسایی وب‌سایت(های) هدف، بازیابی محتوا، استخراج و پردازش داده‌ها، سازمان‌دهی و تحلیل داده‌هاست. پژوهشگر باید با پیروی از قوانین حفاظت از داده و شرایط استفاده از خدمات وب‌سایت‌ها از اخلاقی بودن استخراج اطلاعات، اطمینان حاصل نمایند (Bradley & James, 2019; Stringam et al., 2023). استفاده از وب اسکرپینگ می‌تواند به منظور استخراج داده‌های مورد نیاز برای پژوهش‌های تحلیل استنادی، با استفاده از استخراج استنادها از صفحات نشریات مورد استفاده قرار گیرد. داده‌های متنی مستخرج با استفاده از متن‌کاوی می‌توانند اطلاعاتی بسیار زیادی در اختیار پژوهشگران قرار دهند.

متن‌کاوی که در برگیرنده روش‌های محاسباتی متنوعی مانند پردازش زبان طبیعی، یادگیری ماشین و تحلیل آماری است، به شناسایی واحدهای معنا، روندها و روابط در داده‌های متنی می‌پردازد (Hassani et al., 2020). در حسابداری به دلیل رشد روز افزون وابستگی به داده‌های متنی غیر ساختار یافته، مانند افشاها در گزارشگری مالی، گزارش حسابرسی، و پرونده‌های نظارتی، متن‌کاوی توجه زیادی کسب کرده است (Gupta et al., 2020; Lewis & Young, 2019). بنابراین با توجه به اینکه نوع داده‌ای که از استنادهای مقالات استخراج می‌شود به صورت متن است، متن‌کاوی در تحلیل استنادی می‌تواند بسیار کاربردی باشد. یکی از پرکاربردترین ابزارها، برای سنجش چسبندگی موضوعی، «نمایش رمزگذار دوسویه از تبدیل‌کننده‌های جمله‌ای» است که امکان مقایسه معنایی در سطح جمله، در نتیجه چسبندگی موضوعی را نیز فراهم می‌کند (Dai et al., 2018). بررسی ادبیات نشان می‌دهد که بجای اینکه همه استنادها با یک وزن (تنها شمار استنادها) ملاک ارزیابی واقع شوند، مدل‌های نمایه‌سازی استنادی، هم‌راستایی موضوعی استنادها و موضوع پژوهش را به عنوان وزن برای موزون‌سازی استنادها پیشنهاد می‌کنند (Kazi et al., 2016). هر چند Li et al. (۲۰۲۱) اذعان دارند که نگرانی سوگیری ناشی از الگوریتم همچنان پابرجا خواهد بود. Ranta et al. (۲۰۲۳)، به بررسی نقش یادگیری ماشین در پژوهش‌های حسابداری مدیریت پرداخته‌اند. بررسی جامع پژوهش‌های اخیر نشان می‌دهد که یادگیری ماشین در بهره‌گیری

از منابع متنی گوناگون برای تحلیل داده‌ها، تبدیل داده‌های کیفی به شاخص‌های کمی، بهبود پیش‌بینی‌ها و برآوردها و استفاده از هوش مصنوعی قابل توضیح برای تفسیر مدل‌های یادگیری ماشین است.

Özyigit (۲۰۲۲)، به بررسی کاربردهای نوین متن‌کاوی در حسابداری پرداخته است. این مطالعه نشان می‌دهد که به کارگیری پردازش زبان طبیعی در متن‌کاوی مالی می‌تواند به پیش‌بینی دقیق‌تر روندهای مالی، کاهش خطای انسانی و افزایش کارایی در نظارت بر انطباق مالی منجر شود.

Iqbal et al (۲۰۲۱) در پژوهش مروری خود، ارجاعات درون‌متنی را با استفاده از یادگیری ماشین مورد تحلیل و بررسی قرار دادند. نتایج این پژوهش، اهمیت بکارگیری یادگیری ماشین در تحلیل‌های متنی مانند تحلیل استنادی را آشکار نمود. یادگیری ماشین، دسترسی بهتر به متون کامل مقالات، امکان توسعه روش‌های پیشرفته‌تری مانند دسته‌بندی، تحلیل احساسات، خلاصه‌سازی اسنادها و سیستم‌های پیشنهادگر مبتنی بر استناد را فراهم کرده است. این پیشرفت‌ها عمدتاً با استفاده از تکنیک‌های پردازش زبان طبیعی و یادگیری ماشین محقق شده‌اند.

Dai et al (۲۰۲۱) در پژوهش خود، به تحلیل روند ارجاعات در هشت حوزه مطالعاتی مختلف در بازه زمانی ۱۹۸۰ الی ۲۰۱۹ پرداختند. ایشان دریافتند که در این بازه زمانی، تعداد ارجاعات در هشت حوزه مورد مطالعه روندی افزایشی داشته است.

Song et al (۲۰۲۱)، به بررسی تحلیل کتاب‌سنجی از کاربردهای متن‌کاوی در پژوهش‌های رشته مدیریت پرداختند. نتایج نشان داد که ایالات متحده پیشرو در این حوزه پژوهشی است و بیشترین میزان انتشارات و ارجاعات را دارد. همچنین، موضوعات مدل‌سازی موضوعی، پردازش زبان طبیعی و تحلیل احساسات به عنوان مباحث اصلی در متن‌کاوی در رشته مدیریت شناسایی شده‌اند.

Bertin and Atanassova (۲۰۱۸) به تحلیل و بررسی توزیع چند ارجاعی درون‌متنی پرداختند. نتایج بیانگر آن بود که جملات حاوی چندین استناد درون‌متنی در تمام بخش‌های ساختار مقاله بسیار رایج هستند، با این وجود بخش مقدمه پژوهش‌ها عموماً دارای بیشترین میزان چند استنادی هستند؛ به طوری که بیش از نیمی از جملات حاوی استناد، چندگانه هستند.

Ahmad and Afzal (۲۰۱۸) در پژوهش خود به ارائه الگوریتمی جهت شناسایی لنگر استاندارد پرداختند. نتایج نشان داد که الگوریتم پیشنهادی استاندارد برای شناسایی نشانگرهای استاندارد در متن، با بهبود ۴۱ درصدی میانگین امتیاز F نسبت به روش‌های موجود، عملکرد بهتری دارد.

Fox et al (۲۰۱۶) در پژوهش خود، عوامل گوناگون موثر بر ارجاع‌گیری را مورد بررسی قرار دادند. ایشان دریافتند که عواملی از جمله تعداد رفرنس‌ها، تعداد نویسندگان و حجم پژوهش بیشترین تاثیر مثبت و معنادار را بر ارجاع‌گیری دارند.

Ucar et al (۲۰۱۴) به بررسی روند رشد ارجاعات در پژوهش‌های حوزه مهندسی پرداختند. ایشان در این پژوهش نشان دادند که ارجاعات حوزه‌های مهندسی از سال ۱۹۷۲ تا ۲۰۱۳ افزایشی قابل توجه داشته است. تعداد ارجاعات یکی از معیارهای سنجش علمی و آکادمیک بودن پژوهش‌هاست.

Chan et al (۲۰۰۹) به بررسی رتبه‌بندی مجلات حسابداری بر اساس تحلیل ارجاعات رساله‌های دکتری پرداختند. نتایج نشان داد که چهار مجله برتر حسابداری شامل AR، AOS، JAR و JAE هستند. همچنین، رتبه‌بندی مجلات بسته به حوزه تخصصی و روش پژوهش متفاوت است. به‌عنوان مثال، نشریه AJPT در حوزه حسابداری و BRIA در حوزه تحقیقات رفتاری جایگاه بالاتری دارند. یافته‌ها نشان می‌دهد که رتبه‌بندی‌های سنتی به دلیل سوگیری ممکن است تصویر جامعی ارائه ندهند.

Biglu (۲۰۰۸) تاثیر ارجاعات را بر ضریب تاثیر و اثر متیو مورد بررسی قرار دادند. نتایج پژوهش حاکی از آن بود که تعداد ارجاعات در هر مقاله از ۸,۴۰ در سال ۱۹۷۰ به ۳۴,۶۳ در سال ۲۰۰۵ افزایش یافته. هرچه مجله ای بیشتر به دیگران استناد کند، بیشتر نیز توسط دیگران استناد می‌شود. که با اثر متیو قابل توضیح است.

Mensah et al (۲۰۰۴)، به بررسی میزان تاثیر پژوهش‌های حسابداری مدیریت بر سایر رشته‌های علمی از طریق تحلیل استنادی پرداخته‌اند. یافته‌ها نشان می‌دهد که پژوهش‌های حسابداری مدیریت نه تنها در نشریات حسابداری، بلکه در سایر رشته‌های علمی نیز استناد شده‌اند؛ با این وجود میزان ارجاع به مقالات حسابداری مدیریت در نشریات غیر حسابداری رو به کاهش است که می‌تواند هشدار زودهنگام درباره کاهش ارتباط این حوزه با سایر رشته‌های علمی باشد.

Abt and Garfield (۲۰۰۲) در پژوهش خود به بررسی تعداد رفرنس‌ها و حجم پژوهش در علوم مختلف پرداختند. نتایج این پژوهش نشان داد که مقالات با طول متوسط در علوم گوناگون، تعداد ارجاعات مشابهی دارند. در مقابل، مقالات مروری به طور میانگین دو برابر مقالات تحقیقاتی هم طول خود ارجاعات دارند.

بنابراین با توجه به مطالب ارائه شده، پرسش‌های پژوهش بدین شرح هستند:

- (۱) با توجه به اینکه این تمایل وجود دارد که افزایش تعداد استنادات منتج به استنادات کم اعتبار شوند، اعتبار استنادات انجام شده در پژوهش‌های داخلی چه میزان است؟
- (۲) موضوع استنادات خارجی پژوهش‌های داخلی، تا چه اندازه با موضوع محوری پژوهش هم‌راستا هستند؟

روش

این پژوهش، در چارچوب یک مطالعه کمی و کاربردی با رهیافت تحلیل محتوا و متن کاوی انجام شده است؛ رهیافتی که به گفته Neuendorf (۲۰۱۸)، ابزار دقیقی برای آشکار کردن الگوهای پنهان در داده‌های متنی فراهم می‌آورد. جامعه آماری این پژوهش، تمامی ارجاعات مقالات منتشر شده در نشریات علمی پژوهشی کشور طی سال‌های ۱۳۹۳ الی ۱۴۰۲ را شامل می‌شود. همچنین نمونه آماری پژوهش حاضر، استنادهای خارجی مقالاتی است که در نشریات وزارت علوم، تحقیقات و فناوری حوزه حسابداری و حسابرسی ایران در بازه زمانی ۱۰ ساله (۱۳۹۳ الی ۱۴۰۲) توسط پژوهشگران این حوزه نگاشته شده‌اند و اطلاعات اصلی آن‌ها (نام نویسنده(ها)، عنوان، سال انتشار و استنادهای مقاله) در تارنمای نشریات به صورت متنی وجود دارد. کمیسیون نشریات وزارت علوم، تحقیقات و فناوری ۲۰ مجله حوزه حسابداری را تا پایان سال ۱۴۰۲، واجد رتبه علمی تشخیص داده است. با این حال، تنها در ۱۴ نشریه، متن کامل مقالات به صورت ماشینی قابل استخراج بود و بدین ترتیب این ۱۴ نشریه به‌عنوان چارچوب نمونه انتخاب شدند.

گردآوری داده‌ها با روش خزش وب و با زبان پایتون انجام شد (Khder, 2021). کدنوشت طراحی شده با استفاده از کتابخانه‌های ریکوئستس^۱ و بیوتیفول سوپ^۲ صفحات آرشیو نشریات را پیمایش کرد و برای هر مقاله نام‌برده را ثبت نمود. در نتیجه از مجموع ۳،۴۳۶

^۱ Requests
^۲ BeautifulSoup

مقاله منتشر شده در این نشریات، ۲،۹۶۶ مقاله به دلیل در دسترس بودن داده متنی وارد فرآیند گردآوری شدند. جدول ۱ نمایانگر فرایند گردآوری داده‌هاست.

جدول ۱: تعداد مقالات چاپ شده در نشریات منتخب به همراه تعداد مقالات استخراج شده از تارنمای آن‌ها

تعداد مقالات استخراج شده	تعداد مقالاتی که فاقد اطلاعات متنی در سایت می باشند	تعداد مقالات چاپ شده در دوره استخراج	رتبه ارزیابی (۱۴۰۳)	عنوان نشریه
۲۵۹	۲۱	۲۸۰	الف	بررسی‌های حسابداری و حسابرسی
۴۳۵	۰	۴۳۵	الف	پژوهش‌های تجربی حسابداری
۲۳۱	۱	۲۳۲	الف	پژوهش‌های حسابداری مالی
۱۵۴	۴۲	۱۹۶	الف	پیشرفت‌های حسابداری
۱۶۸	۴	۱۷۲	الف	حسابداری ارزشی و رفتاری
۱۸۸	۰	۱۸۸	الف	حسابداری دولتی
۲۶۶	۳۷	۳۰۳	الف	حسابداری و منافع اجتماعی
۲۶۴	۴۴	۳۰۸	الف	دانش حسابداری
۱۵۵	۱۳۷	۲۹۲	الف	دانش حسابداری مالی
۸۱	۹۴	۱۷۵	الف	Iranian Journal of Auditing, Accounting & Finance
۳۲۸	۳۹	۳۶۷	ب	پژوهش‌های حسابداری مالی و حسابرسی
۸۸	۰	۸۸	ب	پژوهش‌های حسابرسی حرفه‌ای
۳۰۰	۵۱	۳۵۱	ب	حسابداری مدیریت
۴۹	۰	۴۹	ب	قضاوت و تصمیم‌گیری در حسابداری
۲،۹۶۶	۴۷۰	۳،۴۳۶		تعداد کل

تفاوت‌ها در تعداد انتشارات به عوامل گوناگونی، از جمله دوره فعالیت نشریه وابسته است. از متن این مقالات ۱۳۲،۵۴۳ ارجاع بازیابی شد. برای هم‌سویی با هدف پژوهش، یعنی

ارزیابی اثر منابع بین‌المللی، ارجاعات فارسی، کنفرانسی، کتاب‌ها و پایان‌نامه‌ها کنار گذاشته شد (Benson et al., 2014; Moed, 2006; Salimi, 2017; Steinberger, 2010)؛ بنابراین پس از حذف ۶۷،۳۹۷ مورد نامناسب، ۶۵،۱۴۶ ارجاع به مقالات مجلات خارجی باقی ماند. سپس، مقالاتی که پس از پالایش، ارجاع معتبری نداشتند حذف شدند و مجموعه نهایی به ۲،۹۳۸ مقاله تقلیل یافت. به‌منظور آماده‌سازی داده‌ها، همه حروف ارجاعات به حروف کوچک تبدیل، نشانه‌ها و اعداد زائد زدوده، ایست‌واژه‌ها حذف و سپس ساختار هر ارجاع به چهار بخش نام نویسنده(ها)، عنوان مقاله، عنوان نشریه و سال انتشار تفکیک شد. این یکنواخت‌سازی و تفکیک ساختار هر ارجاع با ماکروهای وی‌بی‌ای و کدنوشت پایتون انجام شد تا داده‌ها برای تحلیل‌های بعدی استاندارد گردد.

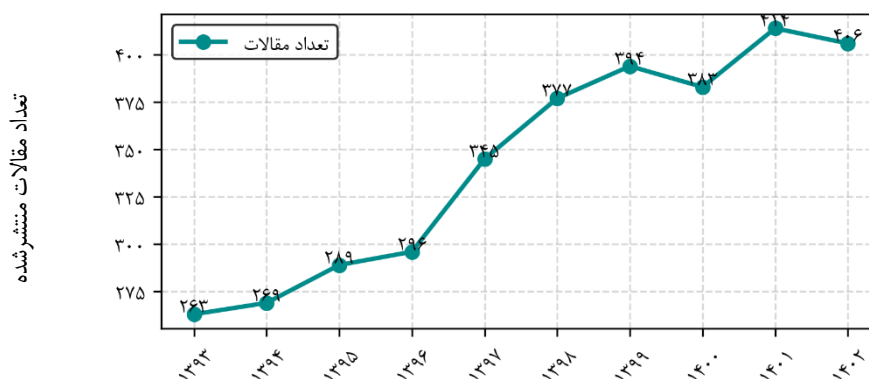
سنجش چسبندگی موضوعی از روش‌های پردازش زبان طبیعی و تحلیل شباهت موضوعی بهره گرفته می‌شود تا میزان شباهت میان عنوان مقالات پژوهشی و استنادهای آنان بررسی شود (Reimers & Gurevych, 2019). ابتدا، کدنوشتی در نرم‌افزار پایتون طراحی شد که عنوان مقالات را با مدل آل‌ام پی‌نت نسخه دو از کتابخانه سنتس ترنسفورمرز به بردارهای عددی تبدیل و سپس شباهت کسینوسی میان این بردارها را محاسبه می‌کند تا به‌صورت خودکار شباهت موضوعی میان هر جفت عنوان (عنوان مقاله اصلی و عنوان استناد) را محاسبه کند. در نهایت، نتایج شباهت‌ها به‌صورت درصد (بین ۰ تا ۱۰۰) در ستون جدیدی در فایل مربوطه ذخیره می‌شود (Song et al., 2020). به استناد پژوهش‌های Reimers and Gurevych (۲۰۱۹)، Feng et al. (۲۰۲۴)، Slater et al. (۲۰۲۲) شباهت‌های کمتر از ۳۰ درصد به‌عنوان شباهت ضعیف، بین ۳۰ تا ۵۰ درصد به‌عنوان شباهت متوسط، بین ۵۰ تا ۷۰ درصد به‌عنوان شباهت خوب و بیش از ۷۰ درصد به‌عنوان شباهت بالا در نظر گرفته شد تا بیشترین همراستایی با پژوهش‌های پیشین و هدف پژوهش حاضر داشته باشد. یکی از برتری‌های این مدل، توانایی آن در مدیریت متون کوتاه و درک مفاهیم عمیق‌تر در مقایسه با مدل‌های پیشین مانند نمایش‌های رمزگذار دو سویه از ترانسفورماتورها است.

یافته‌ها

روند انتشار مقالات در نشریات داخلی

شکل ۱، امکان مشاهده تغییرات سالانه و شناسایی روند کلی را فراهم می‌کند.

شکل ۱: روند انتشار مقالات در نشریات داخلی



رشد پژوهش‌ها از سال ۱۳۹۳ (مقاله ۲۶۳) تا سال ۱۳۹۵ (مقاله ۲۸۹) ممکن است به افزایش تعداد شماره‌های منتشر شده یا ورود نشریات جدید مانند «حسابداری ارزشی و رفتاری» و «Iranian Journal of Accounting, Auditing & Finance» مربوط باشد. از سال ۱۳۹۶، روند انتشار مقالات شتاب بیشتری گرفت اوج انتشار در سال ۱۴۰۱ با ۴۱۴ مقاله مشاهده شد، اما در سال ۱۴۰۲ با کاهشی اندک به ۴۰۶ مقاله رسید. به طور کلی، تعداد مقالات از ۲۶۳ در سال ۱۳۹۳ به ۴۰۶ در سال ۱۴۰۲ افزایش یافت که رشدی ۵۴٫۴ درصدی را نشان می‌دهد. این روند صعودی، نشانه توسعه فعالیت‌های پژوهشی در حوزه حسابداری و حسابرسی در ایران است، هرچند نوسان‌های سال‌های ۱۴۰۰ و ۱۴۰۲ نیاز به بررسی عوامل اثرگذار مانند سیاست‌های نشریات یا اشباع موضوعات پژوهشی را برجسته می‌کند.

تحلیل استناد به نشریات خارجی

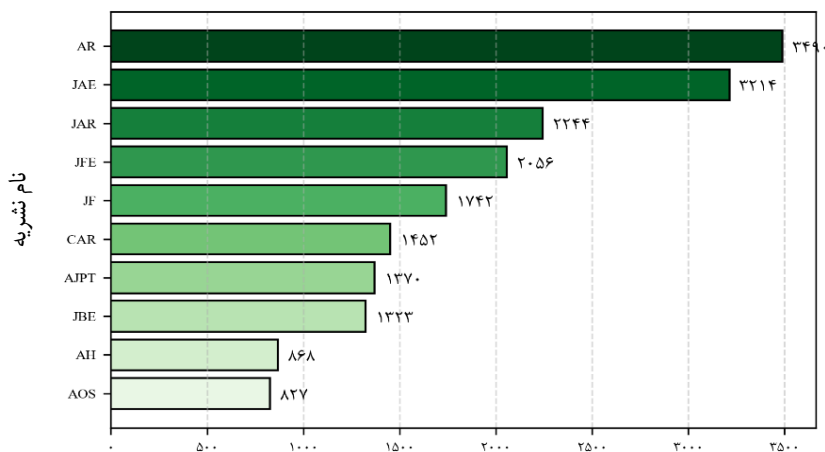
نشریات خارجی پر استناد

برای شناسایی نشریات خارجی پر استناد، داده‌های ۶۵،۱۴۶ استناد تحلیل شد. پس از پاک‌سازی و یکسان‌سازی نام نشریات با استفاده از روش‌های پردازش زبان طبیعی، فراوانی استنادها محاسبه شد. جدول ۲ نام اختصار و کامل ۱۰ نشریه پر استناد را نشان می‌دهد. همچنین شکل ۲ تعداد استنادها به این نشریات را به تصویر می‌کشد. سه نشریه نخست، که در چارک اول رتبه‌بندی SJR قرار دارند، نشان‌دهنده تمرکز پژوهشگران ایرانی بر منابع معتبر و تخصصی حسابداری است.

جدول ۲: نام اختصار ده نشریه پراستناد خارجی

رتبه در ارزیابی سال ۲۰۲۳ SJR	درصد از کل نمونه	تعداد استناد به نشریه	کشور ناشر	نام اختصار نشریه	نام کامل نشریه
۳۳	۵/۴	۳,۴۹۰	ایالات متحده	AR	Accounting Review
۷	۴/۹	۳,۲۱۴	هلند	JAE	Journal of Accounting and Economics
۱۵	۳/۴	۲,۲۴۴	انگلیس	JAR	Journal of Accounting Research
۴	۳/۲	۲,۰۵۶	هلند	JFE	Journal of Financial Economics
۱	۲/۷	۱,۷۴۲	انگلیس	JF	Journal of Finance
۷۴	۲/۲	۱,۴۵۲	کانادا	CAR	Contemporary Accounting Research
۱۶۶	۲/۱	۱,۳۷۰	ایالات متحده	AJPT	Auditing, Journal of Practice and Theory
۹۹	۲/۰	۱,۳۲۳	هلند	JBE	Journal of Business Ethics
۲۹۷	۱/۳	۸۶۸	ایالات متحده	AH	Accounting Horizons
۱۲۸	۱/۳	۸۲۷	انگلیس	AOS	Accounting Organizations and Society
-	۲۸/۵	۱۸,۵۱۶	جمع کل		

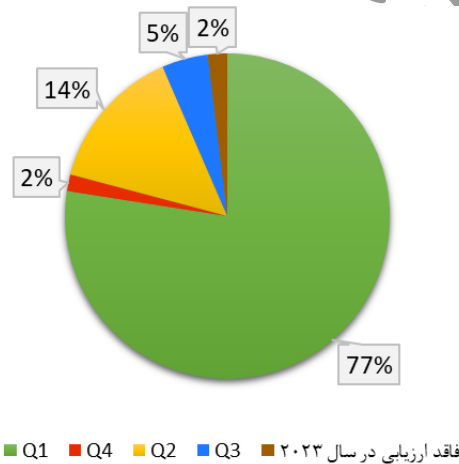
شکل ۲: فهرست ده نشریه پراستناد خارجی



توزیع چارک‌های مورد استناد

توزیع چارک‌های مورد استناد نشان می‌دهد که ۴۱،۱۸۹ استناد به نشریات چارک اول، ۱۷،۶۳۱ استناد به نشریات چارک دوم، ۲،۴۳۴ استناد به نشریات چارک سوم، و ۸۷۳ استناد به نشریات چارک چهارم بوده است. ۱،۰۳۴ استناد نیز به نشریاتی بوده است که در رتبه‌بندی SJR در سال ۲۰۲۳ حضور ندارند، که احتمالاً به دلیل توقف انتشار یا نمایه نشدن در پایگاه اسکوپوس است. شکل ۳ توزیع چارک‌های مورد استناد را مصور می‌سازد.

شکل ۳: توزیع چارک‌های مورد استناد



توزیع چارک نشریات استناد شده در هر نشریه داخلی

جدول ۳: توزیع چارک نشریات استناد شده به تفکیک نشریات داخلی

نام نشریه	چارک نخست	چارک دوم	چارک سوم	چارک چهارم	نشریات فاقد ارزیابی در ۲۰۲۳	نشریات محلی و ناآشنا
پژوهش‌های حسابداری مالی	۷۱/۴	۸/۶	۲/۸	۱/۰	۱/۴	۱۴/۸
مجله دانش حسابداری	۷۰/۷	۹/۳	۳/۱	۱/۳	۱/۴	۱۴/۱
پیشرفتهای حسابداری	۶۷/۶	۱۰/۴	۲/۶	۱/۴	۱/۶	۱۶/۴
بررسی‌های حسابداری و حسابرسی	۶۷/۱	۱۱/۱	۳/۱	۱/۱	۱/۳	۱۶/۳
فصلنامه قضاوت و تصمیم‌گیری در حسابداری	۶۶/۷	۱۰/۶	۴/۰	۱/۱	۱/۳	۱۶/۴
پژوهش‌های تجربی حسابداری	۶۶/۰	۱۰/۸	۲/۷	۱/۲	۱/۶	۱۷/۷
پژوهش‌های حسابداری مالی و حسابرسی	۶۵/۸	۸/۹	۲/۷	۱/۲	۱/۹	۱۹/۵
دانش حسابداری مالی	۶۴/۱	۱۰/۹	۳/۵	۱/۲	۱/۹	۱۸/۳
حسابداری ارزشی و رفتاری	۶۳/۹	۱۰/۳	۳/۰	۱/۳	۱/۷	۱۹/۹
حسابداری مدیریت	۶۲/۰	۱۱/۲	۳/۱	۱/۱	۱/۴	۲۱/۲
حسابداری و منافع اجتماعی	۶۱/۸	۱۱/۴	۳/۳	۱/۳	۱/۹	۲۰/۲

۶۰/۸	۱۱/۷	۴/۴	۱/۲	۱/۷	۲۰/۲	مجله حسابداری، حسابرسی و مالی ایران
۵۸/۷	۱۲/۸	۴/۸	۱/۴	۱/۹	۲۰/۵	پژوهش های حسابرسی حرفه ای
۵۳/۷	۹/۹	۳/۳	۱/۳	۱/۵	۳۰/۳	حسابداری دولتی

جدول ۳ نشان می دهد که ارجاع تعداد قابل توجهی از نشریات ایرانی به نشریات چارک نخست است. این روند حاکی از دقت علمی بالا و تلاش برای کسب اعتبار جهانی است. در مقابل، نشریاتی همچون حسابداری دولتی با ۵۳٫۶۸ درصد، استناد کمتری به نشریات چارک نخست دارند که بیانگر تمرکز بیشتر بر نشریات محلی است. بنابراین، در پاسخ به پرسش نخست پژوهش، اعتبار استنادات در پژوهش های حسابداری و حسابرسی داخلی، در سطح مطلوب ارزیابی می گردد.

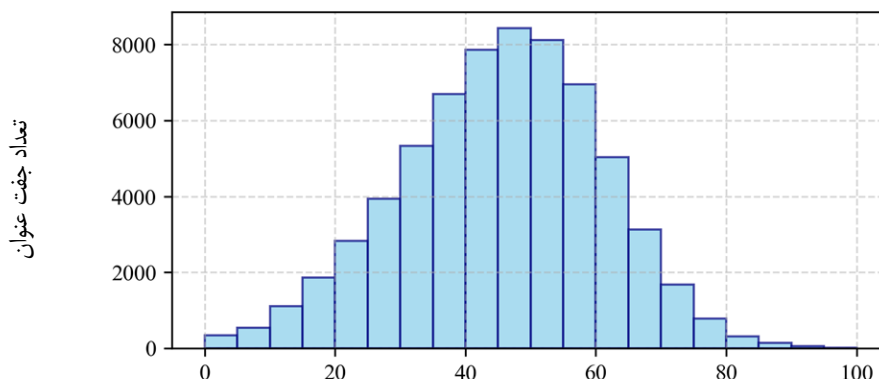
با این حال، استناد به نشریات محلی و ناآشنا با میانگین ۱۸٫۸۳ درصد قابل توجه است و در نشریه حسابداری دولتی به ۳۰٫۳۴ درصد می رسد. این استناددهی ممکن است ناشی از حوزه تخصصی نشریه حسابداری دولتی و یا ترجیحات زبانی باشد. اگرچه این استنادها برای حفظ ارتباط با زمینه های بومی ارزشمند هستند، وابستگی بیش از حد به آنها می تواند جذابیت جهانی نشریات را کاهش دهد.

شباهت موضوعی میان مقالات ارجاعی و مقالات پژوهشگران حسابداری در ایران

شباهت موضوعی ۶۵٫۱۴۶ جفت موضوع (موضوع عنوان پژوهش و موضوع استناد) مورد بررسی قرار گرفت. نتایج این تحلیل در شکل ۴ نمایش داده شده است.

شکل ۴: نمودار هیستوگرام توزیع شباهت موضوعی

درصد شباهت موضوعی



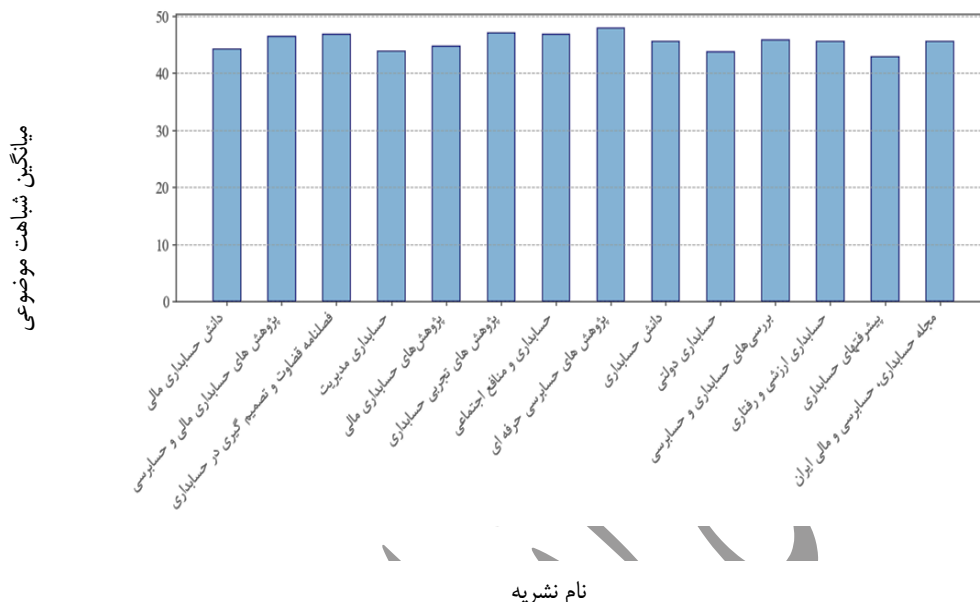
توزیع داده‌ها در هیستوگرام به صورت یک توزیع نرمال با شکل زنگوله‌ای مشاهده می‌شود که اوج آن در بازه‌ای نزدیک به ۴۵ تا ۵۰ درصد قرار دارد. این اوج با میانگین گزارش شده (۴۵٫۴ درصد) هم‌راستا است و نشان می‌دهد که بیشترین تعداد جفت عنوان در این بازه از شباهت موضوعی جای گرفته‌اند. این میزان، در بازه شباهت موضوعی خوب قرار می‌گیرد، اما بسیار نزدیک به مرز شباهت موضوعی متوسط است.

با بررسی دقیق‌تر، می‌توان توزیع داده‌ها را بر پایه بازه‌های تعریف‌شده تحلیل کرد. در بازه شباهت موضوعی ضعیف، فراوانی نسبتاً پایین است و محاسبه می‌شود که ۶،۶۷۸ جفت عنوان (۱۰ درصد از کل داده‌ها) در این بازه قرار دارند. این امر نشان می‌دهد که درصد اندکی از استنادها از نظر معنایی پیوند چندانی با مقاله پژوهشی ندارند و احتمالاً به موضوعات کاملاً متفاوتی اشاره دارند. در بازه شباهت موضوعی بالا تا عالی، فراوانی به تدریج کاهش می‌یابد و به سمت ۱۰۰ درصد بسیار اندک می‌شود. محاسبه می‌شود که ۹ درصد از جفت عنوان‌ها (۶،۱۲۳ جفت عنوان) در این بازه باشند، که نشان‌دهنده آن است که درصد کمی از استنادها پیوند معنایی بسیار نیرومندی با مقاله پژوهشی دارند.

میانگین شباهت موضوعی به تفکیک نشریات داخلی

شکل ۵، میانگین شباهت موضوعی را بر اساس نشریات گوناگون نشان می‌دهد و امکان شناسایی نشریاتی با شباهت موضوعی بالاتر را فراهم می‌کند.

شکل ۵: میانگین شباهت موضوعی در نشریات داخلی

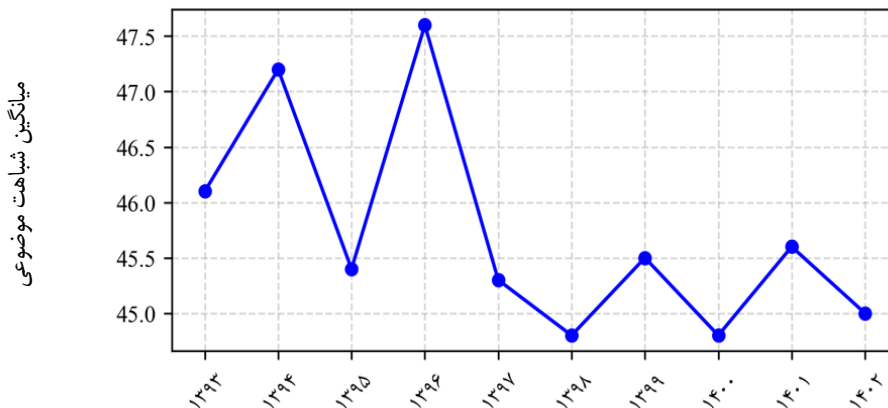


بر اساس نتایج به دست آمده، میانگین شباهت موضوعی در نشریات داخلی بین ۴۲,۸۲ درصد تا ۴۷,۹۳ درصد متغیر است. بالاترین میانگین شباهت به نشریات «پژوهش های حسابرسی حرفه ای» (میانگین ۴۷,۹۳ درصد)، «پژوهش های تجربی حسابداری» (میانگین ۴۶,۹۸ درصد) و «حسابداری و منافع اجتماعی» (میانگین ۴۶,۷۵ درصد) تعلق دارد. بنابراین، در پاسخ به پرسش دوم پژوهش، بیشتر نشریات (۹ نشریه از ۱۴ نشریه) در بازه شباهت موضوعی خوب و ۵ نشریه در بازه متوسط قرار دارند. نشریه «پیشرفت های حسابداری» با میانگین ۴۲,۸۲ درصد کمترین شباهت موضوعی را نشان می دهد.

روند شباهت موضوعی در بازه زمانی مورد مطالعه

میانگین شباهت موضوعی بر اساس سال انتشار مقالات نیز محاسبه شده است. نمودار خطی مصور شده در شکل ۶، میانگین شباهت موضوعی بر اساس سال انتشار مقاله از سال ۱۳۹۳ الی ۱۴۰۲ را نشان می دهد.

شکل ۶: روند میانگین شباهت موضوعی در بازه زمانی مورد مطالعه



میانگین شباهت موضوعی در این دوره زمانی نوسان‌های قابل توجهی داشته است. بالاترین مقدار در سال ۱۳۹۶ (۴۷,۶۱ درصد) و پایین‌ترین مقدار در سال ۱۴۰۰ (۴۴,۷۵ درصد) مشاهده شده است. در بیشتر سال‌ها (۱۳۹۳، ۱۳۹۴، ۱۳۹۵، ۱۳۹۶، ۱۳۹۷، ۱۳۹۹، ۱۴۰۱ و ۱۴۰۲)، میانگین شباهت موضوعی در بازه خوب جای داشته است؛ اما این پیوند معمولاً در سطح متوسط رو به خوب است و کمتر به سطح شباهت بالا رسیده است.

بحث و نتیجه‌گیری

در این پژوهش، با استفاده از رویکرد متن کاوی به تحلیل ارجاعات پژوهش‌های علمی در حوزه‌های حسابداری و حسابرسی پرداخته شد تا با تحلیل روند استنادهای صورت گرفته، تصویری جامع از تأثیر پژوهش‌های خارجی بر پژوهش‌های داخلی ارائه گردد.

نتایج نشان داد که مجلات *Journal of accounting and Accounting Review* و *Journal of accounting research and economics* بیشترین استنادها را از میان نشریه‌های خارجی، در پژوهش‌های داخلی کسب نموده‌اند. همچنین بر اساس داده‌های ارائه شده، ۷۷,۵ درصد از استنادها به نشریات چارک اول (Q1) اختصاص دارد، که بیانگر تمایل پژوهشگران ایرانی به ارجاع به منابع معتبر و با تأثیر بالا است. با توجه به درصد استناد به منابع معتبر، اعتبار کلی استنادها در پژوهش‌های داخلی، مطلوب ارزیابی می‌گردد. نتایج حاصله، برخلاف نظر (Iqbal et al., 2021) بودند که به تمایل به استنادات کم ارتباط به موضوع پژوهش اشاره کردند، چرا که سطح کلی شباهت موضوعی استنادات مطلوب شناخته شدند،

در پژوهش‌های داخلی وجود نداشت. نتایج مطابق با اظهارات de Villiers and Hsiao (۲۰۱۸) درخصوص تمایل پژوهشگران حسابداری به استفاده از نشریات نامبرده است. همچنین چسبندگی (شباهت) موضوعی، «خوب» ارزیابی شد. نتایج نشان می‌دهد که پژوهشگران ایرانی تمایل دارند منابعی را انتخاب کنند که تا حد زیادی با پژوهش‌هایشان هم‌راستا باشد، اما این هم‌راستایی معمولاً در سطح متوسط تا خوب است و در مواردی که منابع عمومی‌تر یا میان رشته‌ای انتخاب می‌شوند، پیوند معنایی کمتری مشاهده می‌شود. این تحلیل بر اهمیت دقت بیشتر در انتخاب منابع مرتبط با موضوعات پژوهش تأکید می‌کند و به پژوهشگران توصیه می‌شود که در انتخاب ارجاعات خود از منابع با شباهت موضوعی بالا استفاده کنند تا انسجام علمی پژوهش‌ها تقویت شود. هرچند نشریاتی نیز شناسایی شدند که چسبندگی موضوعی در آن‌ها کم بود که در این موارد، نتایج مطابق با Iqbal et al (۲۰۲۱) است. در این موارد نیاز است تا داوران سردبیران در ارزیابی مربوط بودن ارجاعات، تلاش بیشتری ورزند. همچنین نتایج این پژوهش، همسو با Anglin et al (۲۰۲۱) نشان داد که می‌توان از مدل‌های مبتنی بر یادگیری ماشین به منظور تحلیل استنادات استفاده نمود.



این پژوهش مانند هر پژوهشی، به‌رغم تلاش در کسب بیشترین دقت در کسب نتایج، با محدودیت‌هایی روبه‌روست. در روش خزیدن وب، پژوهش‌ها از تارنمای نشریات به طور نظام‌مند و خودکار استخراج شدند؛ بنابراین ۴۷۰ پژوهشی که در تارنمای نشریات ارائه نشده‌اند، در پردازش و تحلیل‌ها مورد استفاده قرار نگرفته‌اند که احتمال دارد بر تعمیم‌پذیری نتایج به تمامی پژوهش‌های داخلی مؤثر باشد. با توجه به اینکه در خوشه‌بندی موضوعی از عناوین پژوهش‌های مورد استناد استفاده شده است، پژوهش‌هایی که در عنوان خود از واژگانی با مفاهیم عام و غیر قابل رهگیری به موضوعی از حسابداری یا حسابرسی استفاده کرده‌اند، از داده‌ها کنار گذاشته شده‌اند؛ چراکه واژه یا زنجیره واژگان مورد استفاده در عنوان چنین پژوهش‌هایی، در زمینه پژوهش حاضر، مفهومی خاص از حسابداری یا حسابرسی را منتقل نمی‌کنند.

نتایج این پژوهش در عمل می‌تواند به افزایش کیفیت و اعتبار پژوهش‌های آتی منجر شود. استفاده از رویکردهای مبتنی بر یادگیری ماشین، مانند آنچه در این پژوهش مورد استفاده قرار گرفته است، می‌تواند توسط نشریات داخلی مورد استفاده قرار گیرد تا از مربوط بودن پژوهش‌های مورد ارجاع، اطمینان حاصل کنند. بنابراین با توجه به انگیزه‌هایی که برای

افزایش تعداد ارجاعات برای پژوهشگران وجود دارد، رویکرد اندازه‌گیری چسبندگی موضوعی ارجاعات مورد استفاده در این پژوهش می‌تواند منتج به تعیین آستانه‌ای از چسبندگی موضوعی توسط نشریات شود که نویسندگان را از ارجاعات بی‌ارتباط، وارانند. به طور نمونه، با توجه به اینکه پژوهش‌های معتبر خارجی تأثیر بسزایی بر پیشرفت پژوهش‌های حسابداری داخلی و به دنبال آن انتقال دانش به محیط پژوهشی داخلی را دارند، پیشنهاد می‌شود تا پژوهشگران داخلی، دسترسی فعال‌تری به پایگاه‌های استنادی معتبر مانند وب‌آو‌ساینس و اسکوپوس داشته باشند تا از این طریق دسترسی به پژوهش‌های معتبر و به‌روز تسهیل گردد. همچنین با توجه به دقت ابزارهای نوین در شناسایی الگوهای متنی، پیشنهاد می‌شود تا از پردازش زبان طبیعی و یادگیری ماشین برای شناسایی دقیق‌تر الگوها و شباهت‌ها در داده‌های متنی، توسط پژوهشگران داخلی به میزان بیشتری استفاده شود. افزون بر آن، استفاده از شاخص‌های جدیدتر مانند چسبندگی موضوعی ارجاعات به موضوع اصلی پژوهش می‌تواند ارزیابی مناسب‌تری از اعتبار پژوهش نسبت به توجه صرف به تعداد ارجاعات ارائه دهد؛ بنابراین با توجه به وجود ابزارهای نوظهور برای تحلیل داده‌های متنی، پیشنهاد می‌شود که فرایند بازبینی، داوری و انتخاب پژوهش‌ها برای انتشار، اصلاح یابد و بدین ترتیب شاهد انتشار پژوهش‌هایی بیش از پیش معتبر باشیم.

داده‌های این پژوهش بر روی مجلات داخلی متمرکز است. پژوهش‌های آتی می‌توانند با استفاده از روش و نتایج این پژوهش به سمت بررسی رفتارهای ارجاع‌دهی، مانند منابع و موضوعات مورد ارجاع در کشورهای پیشرفته و همچنین شناسایی محرک‌های این رفتارها مانند تفاوت‌های فرهنگی و جغرافیایی، می‌تواند موضوع پژوهش‌های آتی حرکت کنند.

ORCID

- Hamideh Asnaashari  <https://orcid.org/0000-0003-3648-8899>
- Mohammad Sadra Moradnia  <https://orcid.org/0009-0004-8305-4828>
- Sina Asnaashari  <https://orcid.org/0009-0007-9943-4704>

References

- Abt, H. A., & Garfield, E. (2002). Is the relationship between numbers of references and paper lengths the same for all sciences? *Journal of the American Society for Information Science and Technology*, 53(13), 1106-1112.
<https://doi.org/https://doi.org/10.1002/asi.10151>
- Ahmad, R., & Afzal, M. T. (2018). CAD: an algorithm for citation-anchors detection in research papers. *Scientometrics*, 117(3), 1405-1423. <https://doi.org/10.1007/s11192-018-2920-7>
- Aljuaid, H., Iftikhar, R., Ahmad, S., Asif, M., & Tanvir Afzal, M. (2021). Important citation identification using sentiment analysis of in-text citations. *Telematics and Informatics*, 56, 101492.
<https://doi.org/https://doi.org/10.1016/j.tele.2020.101492>
- Andrew, J., Baker, M., Cooper, C., & Gendron, Y. (2021). Knowledge equity as social justice in academic publishing and why it matters for accounting research. *Meditari Accountancy Research*, 32(6), 2396-2420. <https://doi.org/10.1108/MEDAR-07-2021-0202>
- Anglin, K. L., Wong, V. C., & Boguslav, A. (2021). A Natural Language Processing Approach to Measuring Treatment Adherence and Consistency Using Semantic Similarity. *AERA Open*, 7, 23328584211028615.
<https://doi.org/10.1177/23328584211028615>
- Behrend, J., & Eulerich, M. (2019). The evolution of internal audit research: a bibliometric analysis of published documents (1926–2016). *Accounting History Review*, 29(1), 103-139. <https://doi.org/10.1080/21502851.2019.1676721>
- Benson, K., Faff, R., & Smith, T. (2014). Fifty years of finance research in the Asia Pacific Basin. *Accounting & Finance*, 54(2), 335-373. <https://doi.org/https://doi.org/10.1111/acfi.12081>

- Bertin, M., & Atanassova, I. (2018). The context of multiple in-text references and their signification. *International Journal on Digital Libraries*, 19(2), 127-138.
<https://doi.org/10.1007/s00799-017-0225-7>
- Biglu, M. H. (2008). The influence of references per paper in the SCI to Impact Factors and the Matthew Effect. *Scientometrics*, 74(3), 453-470. <https://doi.org/10.1007/s11192-008-1810-1>
- Boegershausen, J., Datta, H., Borah, A., & Stephen, A. T. (2022). Fields of Gold: Scraping Web Data for Marketing Insights. *Journal of Marketing*, 86(5), 1-20.
<https://doi.org/10.1177/00222429221100750>
- Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45-80.
<https://doi.org/10.1108/00220410810844150>
- Bradley, A., & James, R. J. E. (2019). Web Scraping Using R. *Advances in Methods and Practices in Psychological Science*, 2(3), 264-277. <https://doi.org/10.1177/2515245919859535>
- Chan, K. C., Chan, K. C., Seow, G. S., & Tam, K. (2009). Ranking accounting journals using dissertation citation analysis: A research note. *Accounting, Organizations and Society*, 34(6), 875-885.
<https://doi.org/10.1016/j.aos.2008.12.002>
- Cho, C. H., Jérôme, T., & Maurice, J. (2022). Assessing the impact of environmental accounting research: evidence from citation and journal data. *Sustainability Accounting, Management and Policy Journal*, 13(9), 1014-1030.
<https://doi.org/10.1108/SAMPJ-09-2021-0384>
- Dai, C., Chen, Q., Wan, T., Liu, F., Gong, Y., & Wang, Q. (2021). Literary runaway: Increasingly more references cited per academic research article from 1980 to 2019. *PLOS ONE*, 16(8), e0255849.
<https://doi.org/10.1371/journal.pone.0255849>
- Dai, T., Zhu, L., Cai, X., Pan, S., & Yuan, S. (2018). Explore semantic topics and author communities for citation recommendation in bipartite bibliographic network. *Journal of Ambient*

- Intelligence and Humanized Computing*, 9(4), 957-975.
<https://doi.org/10.1007/s12652-017-0497-1>
- de Villiers, C., & Hsiao, P.-C. K. (2018). A review of accounting research in Australasia. *Accounting & Finance*, 58(4), 993-1026. <https://doi.org/https://doi.org/10.1111/acfi.12424>
- Dechow, P. M., Sloan, R. G., & Zeng, J. (2020). Is It a Home Run? Measuring Relative Citation Rates in Accounting Research. *Accounting Horizons*, 34(1), 67-91.
<https://doi.org/10.2308/acch-52570>
- Endenich, C., & Trapp, R. (2016). Cooperation for Publication? An Analysis of Co-authorship Patterns in Leading Accounting Journals. *European Accounting Review*, 25(3), 613-633.
<https://doi.org/10.1080/09638180.2015.1085886>
- Evans, J. T., Nadjari, H. I., & Burchell, S. A. (1990). Quotational and reference accuracy in surgical journals. A continuing peer review problem. *Jama*, 263(10), 1353-1354.
- Feng, J., Liu, J., Gu, C., Qi, H., Ren, Z., Xu, K., & Wang, Y. (2024). Focusing on differences! Sample framework enhances semantic textual similarity with external knowledge. *Expert Systems with Applications*, 255, 124462.
<https://doi.org/https://doi.org/10.1016/j.eswa.2024.124462>
- Fox, C. W., Paine, C. E. T., & Sauterey, B. (2016). Citations increase with manuscript length, author number, and references cited in ecology journals. *Ecology and Evolution*, 6(21), 7717-7726.
<https://doi.org/https://doi.org/10.1002/ece3.2505>
- García Osma, B., Araceli, M., & Pierk, J. (2023). Dissemination of Accounting Research. *European Accounting Review*, 32(5), 1053-1083.
<https://doi.org/10.1007/s11142-023-10001-1>
- Gupta, A., Dengre, V., Kheruwala, H. A., & Shah, M. (2020). Comprehensive review of text-mining applications in finance. *Financial Innovation*, 6, 1-25.
- Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). (<https://doi.org/10.3390/bdcc4010001>). *Big Data and Cognitive Computing*, 4(1), 1.

- Iqbal, S., Hassan, S.-U., Aljohani, N. R., Alelyani, S., Nawaz, R., & Bornmann, L. (2021). A decade of in-text citation analysis based on natural language processing and machine learning techniques: an overview of empirical studies. *Scientometrics*, 126(8), 6551-6599. <https://doi.org/10.1007/s11192-021-04000-1>
- Jones, S., & Alam, N. (2019). A machine learning analysis of citation impact among selected Pacific Basin journals. *Accounting & Finance*, 59(4), 2509-2552. <https://doi.org/https://doi.org/10.1111/acfi.12584>
- Kazi, P., Patwardhan, M., & Joglekar, P. (2016). Towards a new perspective on context based citation index of research articles. *Scientometrics*, 107(1), 103-121 . <https://doi.org/10.1007/s11192-016-1844-2>
- Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3). <https://doi.org/https://doi.org/10.15849/ijasca.211128.11>
- Khelif, H., & Chalmers, K. (2015). A review of meta-analytic research in accounting. *Journal of Accounting Literature*, 35, 1-27. <https://doi.org/https://doi.org/10.1016/j.acclit.2015.09.001>
- Kim, B., Choi, S.-h., Kang, M & , Kang, J.-H. (2015). Analysis of Research status based on Citation Context. *International Journal of Contents*, 11(2), 63-68. <https://doi.org/http://doi.org/10.5392/IJoC.2015.11.2.063>
- Lewis, C., & Young, S. (2019). Fad or future? Automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research*, 49(5), 587-615. <https://doi.org/https://doi.org/10.1080/00014788.2019.1611473>
- Li, F., Zhou, Y., & Cai, T. (2021). Trails of Data: Three Cases for Collecting Web Information for Social Science Research. *Social Science Computer Review*, 39(5), 922-942. <https://doi.org/10.1177/0894439319886019>
- Mashayekhi, B., Dolatzarei, E., Faraji, O., & Rezaee, Z. (2024). Mapping the state of expanded audit reporting: a bibliometric

- view. *Meditari Accountancy Research*, 32(2), 579-612.
<https://doi.org/10.1108/MEDAR-09-2022-1809> [In Persian]
- Mensah, Y. M., Hwang, N. C. R., & Wu, D. (2004). Does Managerial Accounting Research Contribute to Related Disciplines? An Examination Using Citation Analysis. *Journal of Management Accounting Research*, 16(1), 163-181.
<https://doi.org/10.2308/jmar.2004.16.1.163>
- Meyer, M., Waldkirch, R. W., Duscher, I., & Just, A. (2018). Drivers of citations: An analysis of publications in “top” accounting journals. *Critical Perspectives on Accounting*, 51, 24-46.
<https://doi.org/https://doi.org/10.1016/j.cpa.2017.07.001>
- Moed, H. F. (2006). *Citation analysis in research evaluation* (Vol. 9). Springer Science & Business Media.
https://link.springer.com/book/10.1007/978-1-4020-3371-4_5
- Myers, N., Snow, N., Summers, S. L., & Wood, D. A. (2016). Accounting institution citation-based research rankings by topical area and methodology. *Journal of Information Systems*, 30(3), 33-62.
<https://doi.org/https://doi.org/10.2308/isys-51343>
- Negash, M., Lemma, T. T., & Samkin, G. (2019). Factors impacting accounting research output in developing countries: An exploratory study. *The British Accounting Review*, 51(2), 170-192.
<https://doi.org/https://doi.org/10.1016/j.bar.2018.09.003>
- Neuendorf, K. A. (2018). Content analysis and thematic analysis. In *Advanced research methods for applied psychology* (pp. 211-223). Routledge.
<https://www.taylorfrancis.com/chapters/edit/10.4324/9781315017971-21/content-analysis-thematic-analysis-kimberly-neuendorf>
- Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology*, 66(4), 739-764.
<https://doi.org/https://doi.org/10.1002/asi.23209>
- Özyiğit, H. (2022). (MUHASEBE ALANINA GÜNCEL YAKLAŞIMLAR: METİN MADENCİLİĞİ. *Journal of Accounting and Taxation*

Studies, 15(3), 637-663.

<https://doi.org/https://doi.org/10.29067/muvu.1104525>

Ranta, M., Mika, Y., & Järvenpää, M. (2023). Machine Learning in Management Accounting Research: Literature Review and Pathways for the Future. *European Accounting Review*, 32(3), 707-737. <https://doi.org/10.1080/09639284.2022.2137221>

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

<https://doi.org/https://doi.org/10.48550/arXiv.1908.10084>

Salimi, N. (2017). Quality assessment of scientific outputs using the BWM. *Scientometrics*, 112, 195-213.

<https://doi.org/https://doi.org/10.1007/s11192-017-2284-7>

Sangster, A., Fogarty, T., Stoner, G., & Marriott, N. (2015). The impact of accounting education research. *Accounting Education*, 24(5), 423-444.

<https://doi.org/https://doi.org/10.1080/09639284.2015.1091741>

Shi, H. (2024). Natural Language Processing and Text Mining Algorithms for Financial Accounting Information Disclosure. *Journal of Electrical Systems*, 20(9s), 453-461.

<https://journal.esrgroups.org/jes/article/view/4346>

Silvello, G. (2018). Theory and practice of data citation. *Journal of the Association for Information Science and Technology*, 69(1), 6-20. <https://doi.org/https://doi.org/10.1002/asi.23917>

Skolarus, T. A., Lehmann, T., Tabak, R. G., Harris, J., Lecy, J., & Sales, A. E. (2017). Assessing citation networks for dissemination and implementation research frameworks. *Implementation Science*, 12(1), 97. <https://doi.org/10.1186/s13012-017-0628-7>

Slater, L. T., Russell, S., Makepeace, S., Carberry, A., Karwath, A., Williams, J. A., Fanning, H., Ball, S., Hoehndorf, R., & Gkoutos, G. V. (2020). Evaluating semantic similarity methods for comparison of text-derived phenotype profiles. *BMC Medical Informatics and Decision Making*, 22(1), 33.

<https://doi.org/10.1186/s12911-022-01770-4>

- Song, G., Wu, J., & Wang, S. (2021). [Retracted] Text Mining in Management Research: A Bibliometric Analysis. *Security and Communication Networks*, 2021(1), 2270276.
<https://doi.org/https://doi.org/10.1155/2021/2270276>
- Song, K., Tan, X., Qin, T., Lu, J., & Liu, T.-Y. (2020). Mpnet: Masked and permuted pre-training for language understanding. *Advances in neural information processing systems*, 33, 16857-16867.
<https://doi.org/https://doi.org/10.48550/arXiv.2004.09297>
- Steinberger, R. (2010). Challenges and methods for multilingual text mining. Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)
<http://www.lrec-conf.org/proceedings/lrec2010/keynotes/keynote2.pdf> ,
- Stringam, B., Gerdes, J. H., & Anderson, C. K. (2023). Legal and Ethical Issues of Collecting and Using Online Hospitality Data. *Cornell Hospitality Quarterly*, 64(1), 54-62.
<https://doi.org/10.1177/19389655211040434>
- Ucar, I., López-Fernandino, F., Rodríguez-Ulibarri, P., Sesma-Sanchez, L., Urrea-Micó, V., & Sevilla, J. (2014). Growth in the number of references in engineering journal papers during the 1972–2013 period. *Scientometrics*, 98(3), 1855-1864.
<https://doi.org/10.1007/s11192-013-1113-6>
- van Raan, A. F., Visser, M. S., Van Leeuwen, T. N., & van Wijk, E. (2003). Bibliometric analysis of psychotherapy research: performance assessment and position in the journal landscape. *Psychother Res*, 13(4), 511-528.
<https://doi.org/10.1093/ptr/kpg038>
- Woolgar, S. (1991). Beyond the citation debate: towards a sociology of measurement technologies and their use in science policy. *Science and Public Policy*, 18(5), 319-326.
<https://doi.org/10.1093/spp/18.5.319>

پیش از انتشار

استناد به این مقاله: نام خانوادگی نویسنده اول، نام. (سال). عنوان مقاله. عنوان نشریه (ایتالیک)، سال (شماره)، ص
آغاز-ص پایان.



Name of Journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

پیش از انتشار