

## ■ ارزیابی زوج‌های کتابشناختی در خوشه‌بندی

### پروانه‌های ثبت اختراع

آناهیتا کرمانی | نرگس نشاط

## ■ چکیده

هدف: خوشه‌بندی زوج‌های کتابشناختی (اشتراک استنادی) یکی از شکل‌های خوشه‌بندی مدارک است. هدف این پژوهش شناسایی میزان جامعیت و مانعیت این نوع خوشه‌بندی در پروانه‌های ثبت اختراع است.

روش/ رویکرد پژوهش: ۷۱۷ پروانه ثبت اختراع یو.اس. مورد استناد در پروانه‌های ثبت اختراع مربوط به رده ۹۷۷/۷۷۴ از رده‌بندی پروانه‌های ثبت اختراع یو.اس. با استفاده از زوج‌های کتابشناختی و الگوریتم فازی خوشه‌بندی شد. حاصل کار با استفاده از معیارهای ارزیابی مانعیت و جامعیت گسترش یافته بی‌کیوبد مورد ارزیابی قرار گرفت. یافته‌ها: خوشه‌بندی پروانه‌های ثبت اختراع با استفاده از اشتراک استنادی عملکرد مطلوبی داشته و از جامعیت گسترده‌ای برخوردار بوده است.

نتیجه‌گیری: با توجه به عملکرد مطلوب خوشه‌بندی بر مبنای زوج‌های کتابشناختی، استفاده از زوج‌های کتابشناختی در فرآیند سازماندهی و ارزیابی پروانه‌های ثبت اختراع نیز می‌تواند عملکرد مطلوبی داشته باشد.

## کلیدواژه‌ها

پروانه‌های ثبت اختراع، زوج‌های کتابشناختی، خوشه‌بندی فازی c- میانگین، خوشه‌بندی، ارزیابی

# ارزیابی زوج‌های کتابشناختی در خوشه‌بندی پروانه‌های ثبت اختراع

آناهیتا کرمانی<sup>۱</sup> | نرگس نشاط<sup>۲</sup>

دریافت: ۱۳۹۰/۷/۴ پذیرش: ۱۳۹۱/۱/۳۰

## مقدمه

پروانه‌های ثبت اختراع یکی از منابع مهم اطلاعاتی در فناوری به‌شمار می‌رود که اطلاعات موجود در آنها می‌تواند در اهداف گوناگون مورد استفاده قرار گیرد. «تعیین قابلیت ثبت» اختراعات مورد ادعا و یا «رد اعتبار» اختراعات ثبت شده به‌واسطه یافتن اختراعی مشابه و یا «پیمایش‌های فناوری» از جمله کارکردهای اطلاعات موجود در پروانه‌های ثبت اختراع است. در این راستا، ایجاد نظام‌های بازیابی مناسب و نیز به‌کارگیری روش‌های مناسب در ایجاد نقشه‌های فناوری نقش بسزایی در استفاده بهینه از اطلاعات موجود این منابع ایفا می‌کند.

گروه‌بندی مدارک مرتبط همواره یکی از دغدغه‌های اصلی در حوزه‌هایی چون سازماندهی و بازیابی اطلاعات بوده است. پروانه‌های ثبت اختراع ممکن است به دلایلی چون بررسی میزان تازگی و غیربیدیگی بودن یک اختراع یا رد اعتبار آن، مورد جست‌وجو قرار گیرند. ارزیاب<sup>۳</sup> مأمور به بررسی ماهوی یک اختراع به دلیل بررسی تازگی و غیربیدیگی بودن آن، و شخصی که مدعی تجاوز به حق اختراع خود است باید برای رد اعتبار اختراع ثبت شده، به جست‌وجو در منابع و از جمله پروانه‌های ثبت اختراع بپردازد (گراف، و آزوپاردی<sup>۴</sup>، ۲۰۰۸؛ اداره ثبت اختراع و علائم تجاری امریکا<sup>۵</sup>، ۲۰۱۰؛ بونینو، شیاراملا، کورنو<sup>۶</sup>، ۲۰۱۰، نقل در هیدئو، آزوپاردی، و ندریاووهده<sup>۷</sup>، ۲۰۱۰).

همه این موارد فقط گوشه‌ای از نیازهایی است که ضرورت و اهمیت سازماندهی و

۱. کارشناس ارشد کتابداری و اطلاع‌رسانی

kermanianahita@gmail.com

(نویسنده مسئول)

۲. دانشیار سازمان اسناد و کتابخانه ملی ایران

narges\_neshat@yahoo.com

3. Examiner

4. Graf & Azzopardi

5. United States Patent and Trademark Office (USPTO)

6. Bonino, Ciaramella, & Como

7. Hideo, Azzopardi,

& Vanderbauwhede

بازیابی مناسب پروانه‌های ثبت اختراع را نشان می‌دهد. در این راستا، پژوهش‌های متفاوتی پیرامون سازماندهی و بازیابی اطلاعات پروانه‌های ثبت اختراع انجام شده که در برخی از آنها (کانگ<sup>۸</sup> و دیگران، ۲۰۰۷؛ بشیر، و راب<sup>۹</sup>، ۲۰۰۹) نیز به روش خوشه‌بندی پرداخته شده است.

خوشه‌بندی روشی است که برای گروه‌بندی موجودیت‌های<sup>۱۱</sup> (مدارک) مشابه مورد استفاده قرار می‌گیرد. در این روش، مدارک در گروه‌هایی از پیش تعیین نشده به نام خوشه قرار می‌گیرند؛ به طوری که مدارک مشابه در کنار یکدیگر و مدارک نامشابه دور از یکدیگر قرار گیرند (تن، اشتینباخ، و کومار<sup>۱۱</sup>، ۲۰۰۶).

عوامل گوناگونی در خوشه‌بندی مؤثر است؛ یکی از این عوامل مربوط به نوع خصیصه‌های<sup>۱۲</sup> است که به واسطه آن یک مدرک مورد بازنمایی قرار می‌گیرد. نوع خصیصه انتخابی به همراه الگوریتم مناسب، برای خوشه‌بندی از اهمیت بسیار بالایی است (دیلن، کگان و نیکلاس<sup>۱۳</sup>، ۲۰۰۳). به این دلیل، انتخاب درست خصیصه برای مدارک در نتایج خوشه‌بندی تأثیرگذار خواهد بود.

غالباً خصیصه‌های گوناگونی برای بازنمایی<sup>۱۴</sup> پروانه‌های ثبت اختراع مورد استفاده قرار می‌گیرد که استناد یکی از پرکاربردترین آنهاست. اهمیت استنادها به این دلیل است که از یک سو توسط افراد متخصصی به نام ارزیاب تعیین می‌شوند و از سوی دیگر، این استنادها مرتبط با ادعاهای موجود در اختراع هستند و اختراعی بودن یک اثر را مورد تأیید قرار می‌دهند<sup>۱۵</sup>. از این رو، می‌توانند به عنوان خصیصه‌ای مناسب برای بازنمون پروانه‌های ثبت اختراع مورد استفاده قرار گیرند.

در میان منابع فارسی و خارجی اثری که در آن خوشه‌بندی پروانه‌های ثبت اختراع را با استفاده از عناصر استنادی مورد ارزیابی قرار داده باشد یافت نشد، اما می‌توان به آثار لای، و وو<sup>۱۶</sup> (۲۰۰۵)، لی<sup>۱۷</sup> و دیگران (۲۰۰۷)، فوجی<sup>۱۸</sup> (۲۰۰۷) اشاره کرد که عنصر استناد را به نوعی در گروه‌بندی پروانه‌های ثبت اختراع مورد توجه قرار داده‌اند.

لای و وو (۲۰۰۵) در مقاله خود با عنوان «استفاده از رویکرد هم استنادی برای ایجاد سیستمی نوین در رده‌بندی پروانه‌های ثبت اختراع»<sup>۱۹</sup> استفاده از رویکرد هم استنادی را به منظور رده‌بندی پروانه‌های ثبت اختراع پیشنهاد کرده‌اند. نتایج ارزیابی این مدل با استفاده از روش تجربی در مجموعه‌ای آزمایشی شامل پروانه‌های مربوط به نانو تکنولوژی، حاکی از مفید بودن اطلاعات استنادی در سازماندهی پروانه‌های ثبت اختراع بوده است.

کانگ و دیگران (۲۰۰۷) در مقاله‌ای با عنوان «بازیابی خوشه-محور برای پروانه‌های ثبت اختراع»<sup>۲۰</sup> به مطالعه بازیابی مبتنی بر خوشه‌بندی پروانه‌های ثبت اختراع پرداخته‌اند که

8. Kang
9. Bashir, & Rauber
10. Entity
11. Tan, Steinbach & Kumar
12. Attribute
13. Dhillon, Kogan & Nicholas
14. Representation
15. البته باید توجه داشت که نظام‌های ثبت اختراع در مراکز گوناگون متفاوت است و توصیف یاد شده از استناد در مراکز صدق می‌کند که در آن ثبت اختراع مستلزم بررسی ماهوی اختراع است.
16. Lai & Wu
17. Li
18. Fujii
19. Using the patent co-citation approach to establish a new patent classification system
20. Cluster-based patent retrieval

در آن هدف از بازیابی، رد اعتبار پروانه‌های ثبت اختراع بوده است. یافته‌ها حاکی از آن بود که مدل زبانی مبتنی بر خوشه می‌تواند مدل زبانی فاقد خوشه را بهبود بخشد.

لی و دیگران (۲۰۰۷) در مقاله‌ای با عنوان «رده‌بندی خودکار پروانه‌های ثبت اختراع با استفاده از شبکه استنادی: مطالعه‌ای تجربی در فناوری نانو»<sup>۲۱</sup>، استفاده از اطلاعات استنادی پروانه‌های ثبت اختراع به‌ویژه ساختار شبکه استنادی را در اختصاص خودکار رده‌ها به پروانه‌های ثبت اختراع مورد بررسی قرار داده‌اند. نتایج این تحقیق حاکی از آن است که استفاده از ساختارهای شبکه استنادی به‌طور معنی‌داری عملکردی بهتر نسبت به حالتی ایجاد می‌کند که در آن از استناد استفاده نشده است.

فوجی (۲۰۰۷) در مقاله‌ای با عنوان «تلفیق اطلاعات استنادی و محتوایی در بازیابی ثبت اختراع ان. تی. سی. آی. آر. ۶»<sup>۲۲</sup> به ارزیابی بازیابی اطلاعات پروانه‌های ثبت اختراع با هدف رد اعتبار پروانه‌های ثبت اختراعات پرداخته است. مدل پیشنهادی او شامل تلفیق اطلاعات استنادی و متنی برای بازیابی پروانه‌های ثبت اختراع بوده است. در این مدل، ابتدا بازیابی مبتنی بر متن و پس از آن، از اطلاعات استنادی استفاده شده است. در بازیابی مبتنی بر متن، متن ادعای (های) موجود در پروانه‌های ثبت اختراع مورد استفاده قرار گرفته است. در نمره‌دهی استنادی مدارک دو شیوه در پیش گرفته شده: حالتی که نمره استنادی در کل مدارک در نظر گرفته می‌شود؛ و حالتی که نمره استنادی در N مدارک برتر حاصل از بازیابی مبتنی بر متن محاسبه می‌شود. در این مدل، به منظور نمره‌دهی ثبت اختراعات از نظر اطلاعات استنادی از روش پیچ رنگ<sup>۲۳</sup> استفاده شده است. نتایج این پژوهش تجربی نشان داده که تلفیق اطلاعات استنادی در حالت اول (محاسبه نمره استنادی بدون در نظر گرفتن N مدارک برتر) اثربخش‌تر از حالت دوم عمل کرده، اما به‌طور کلی، تلفیق اطلاعات استنادی و متنی در بازیابی اثربخش بوده است.

تیوانا، و هورویترز<sup>۲۴</sup> (۲۰۰۹) در «فایندسایت - یافتن خودکار پروانه‌های ثبت اختراع پیشین»<sup>۲۵</sup> به ارائه الگوریتمی پرداخته‌اند که در آن از استنادها برای یافتن پروانه‌های ثبت اختراع استفاده شده است. این الگوریتم بر روی نتایج بازیابی پیاده‌سازی می‌شود. به این معنی که ابتدا واژه‌ای جست و جو و پس از آن الگوریتم بر روی مجموعه مدارک بازیابی شده، پردازش‌های بعدی را انجام می‌دهد. ابتدا از طریق جست‌وجوی کلیدواژه‌ای، مجموعه از پروانه‌های ثبت اختراع مشخص و سپس پروانه‌های ثبت اختراع مهم از طریق استنادهای پیش‌آیند و پس‌آیند مشخص می‌شود. استفاده از استناد در الگوریتم پیشنهادی سبب بازیابی تعداد زیادی از منابع نامرتبط در کنار منابع مرتبط بوده است.

زو، و کرافت<sup>۲۶</sup> (۲۰۰۹) در مقاله خود با عنوان «تبدیل پروانه‌های ثبت اختراع به عبارات‌های جستجو برای آثار پیشین»<sup>۲۷</sup> به شناسایی بهترین ناحیه برای انتخاب خودکار

21. Automatic patent classification using citation network information: An experimental study in nanotechnology

22. Integrating content and citation information for the NTCIR-6 patent retrieval task

23. PageRank

24. Tiwana, & Horwitz

25. Findcite: Automatically finding prior art patents

26. Xuo & Croft

27. Transforming patents into prior-art queries

عبارت جست‌وجو پرداخته‌اند و به‌منظور کاهش مسئولیت کاربران، کل مدرک ثبت اختراع به‌عنوان عبارت جست‌وجو در نظر گرفته شده است. نتایج حاصل از تحقیق تجربی بر روی مجموعه حاوی پروانه‌های ثبت اختراع اداره یو. اس. پی. تی. ا. ۲۸ حاکی از آن است که از میان نواحی عنوان، چکیده، خلاصه آثار پیشین، توصیف اشکال، شرح جزئیات، ادعاها، جست‌وجوی آزاد و ادعاهای مقدم، ناحیه خلاصه آثار پیشین<sup>۲۹</sup> دارای بهترین تأثیر در بازیابی آثار پیشین بوده است. همچنین در میان الگوریتم‌های مورد استفاده برای وزن‌دهی، الگوریتم *tf* دارای بهترین عملکرد و پس از آن، الگوریتم‌های *tf-idf* و *bool* قرار داشته‌اند.

و نگریری و آدریانی<sup>۳۰</sup> (۲۰۱۰) در پژوهش خود با عنوان «بازیابی آثار پیشین با استفاده از محتوای نواحی مختلف در مدارک پروانه ثبت اختراع»<sup>۳۱</sup>، عملکرد بازیابی به‌منظور رد اعتبار<sup>۳۲</sup> یک مدرک ثبت اختراع را هنگام استفاده از سه ناحیه «ادعاها»، ترکیب «عنوان و توصیف» و ترکیب «عنوان، توصیف و ادعاها» مورد بررسی قرار داده‌اند. نتایج نشان داده که در میان نواحی مورد بررسی، ترکیب «عنوان، توصیف و ادعاها» دارای بهترین عملکرد از لحاظ معیارهای جامعیت، مانعیت، مانعیت میانگین متوسط (MAP) و NDCG است. همچنین میزان جامعیت در هر دو مجموعه موضوعات عام<sup>۳۴</sup> و خاص<sup>۳۵</sup> بسیار بیشتر از میزان مانعیت آن بوده، و MAP موضوعات عام (۰/۰۵۸۱) بیشتر از موضوعات خاص (۰/۰۵۳۷) است.

چنانچه مشاهده می‌شود استفاده از استناد به‌عنوان خصیصه می‌تواند به دو شیوه انجام شود: از یک سو، در خوشه‌بندی منابع «استنادکننده» و از سوی دیگر در خوشه‌بندی منابع «استنادشونده». روش نخست که در آن «استنادهای موجود در یک مدرک»، خصیصه‌ای برای بازنمون مدارک «استنادکننده» محسوب شود «زوج‌های کتابشناختی»<sup>۳۶</sup>، و روش دوم که «مدارک استنادکننده» به‌عنوان خصیصه‌ای برای مدارک «استناد شونده» محسوب شود «هم‌استنادی»<sup>۳۷</sup> نام دارد. از آنجا که در پژوهش حاضر خوشه‌بندی منابع استنادکننده مورد نظر بوده است، لذا بر آنیم تا عملکرد زوج‌های کتابشناختی را در خوشه‌بندی پروانه‌های ثبت اختراع مورد بررسی قرار داده و برای پرسش‌های زیر پاسخی مناسب بیابیم:

۱. میزان جامعیت در خوشه‌بندی پروانه‌های ثبت اختراع با استفاده از زوج‌های کتابشناختی چقدر است؟
۲. میزان مانعیت در خوشه‌بندی پروانه‌های ثبت اختراع با استفاده از زوج‌های کتابشناختی چقدر است؟

## روش پژوهش و جامعه مورد مطالعه

پژوهش حاضر به روش تجربی انجام شده و جامعه آماری شامل ۷۱۷ پروانه ثبت اختراع

28. USPTO
29. Background summary
30. Wangiri & Adriani
31. Prior art retrieval using various patent document fields contents
32. Invalidate
33. Mean Average Precision
34. Large topic set
35. Small topic set
36. Bibliographic coupling
37. Co-citation

یو. اس.<sup>۳۸</sup> مورد استناد در پروانه‌های ثبت اختراع یو. اس. متعلق به رده ۹۷۷/۷۷۴ از رده‌بندی یو. اس. پی. سی.<sup>۳۹</sup> است. رده ۹۷۷ رده‌ای اصلی و مربوط به موضوع نانو تکنولوژی (فناوری نانو)؛ رده ۹۷۷/۷۷۴ رده‌ای فرعی و زیرمجموعه‌ای<sup>۴۰</sup> از رده فرعی ۹۷۷/۷۷۳؛ رده ۹۷۷/۷۷۳ نیز زیرمجموعه‌ای از رده ۹۷۷/۷۰۰ است. رده فرعی ۹۷۷/۷۰۰ مربوط به نانو ساختارها و رده فرعی ۹۷۷/۷۷۳ شامل نانو ذرات است.<sup>۴۱</sup>

پس از استخراج و ذخیره اطلاعات در نرم‌افزار اکسل، ابزارهای زیر برای مراحل مختلف مورد استفاده قرار گرفت:

نرم‌افزار کد منبع باز ریپید ماینر<sup>۴۲</sup> در وزن‌دهی به اسنادها؛ برنامه<sup>۴۳</sup> خوشه‌بندی فازی<sup>۴۴</sup> نوشته‌شده به زبان برنامه‌نویسی پرل<sup>۴۵</sup>؛ و برنامه<sup>۴۶</sup> محاسباتی جامعیت و مانعیت گسترش یافته بی کیوبد که در گارگاه وبس<sup>۴۷</sup> معرفی و ارائه شده است. در اجرای پژوهش، فرآیندهایی از جمله ایجاد مجموعه مورد آزمایش<sup>۴۸</sup>، استخراج اسنادها، ایجاد ماتریس‌های استناد - پروانه ثبت اختراع، خوشه‌بندی پروانه‌های ثبت اختراع انجام شده است که در ادامه به آنها می‌پردازیم.

## نحوه ایجاد مجموعه آزمایشی و استخراج اسنادها

ارزیابی خوشه‌بندی معمولاً با استفاده از مجموعه‌های آزمایشی انجام می‌شود. مجموعه‌های آزمایشی دربرگیرنده مجموعه مدارکی است که نسبت به آنها قضاوت‌های ربط صورت گرفته است. در این مجموعه‌ها موضوعاتی تعریف شده که تعدادی از مدارک موجود در مجموعه آزمایشی، به آنها مرتبط است. در این حالت ارزیابی شیوه پیشنهادی برای خوشه‌بندی، با توجه به قابلیت آن در گروه‌بندی مدارک مرتبط، مورد ارزیابی قرار می‌گیرد. یکی از روش‌های ایجاد مجموعه‌های آزمایشی برای پروانه‌های ثبت اختراع، ایجاد مجموعه‌های آزمایشی مبتنی بر استناد است. در این حالت هر پروانه ثبت اختراع مانند موضوعی است که اسنادهای موجود در آن، مدارک مرتبط به آن موضوع را مشخص می‌کند. در پژوهش حاضر نیز چنین رویکردی در ایجاد مجموعه‌های آزمایشی مورد استفاده قرار گرفته است. بدین صورت که مجموعه پروانه‌های ثبت اختراع متعلق به رده‌ای خاص به‌عنوان موضوع و اسنادهای موجود در آن به‌عنوان مدارک مرتبط با آن موضوع در نظر گرفته شده است. بنابراین مجموعه مورد آزمایش در تحقیق حاضر مجموعه مدارکی خواهد بود که پروانه‌های ثبت اختراع رده مفروض آن را مورد استناد قرار داده باشند.

برای ایجاد مجموعه مورد آزمایش ابتدا تعداد ۱۰۴ پروانه ثبت اختراع مربوط به رده ۹۷۷/۷۷۴ به‌عنوان خوشه‌بندی مرجع برگزیده شد. برای به‌دست آوردن پروانه‌های ثبت اختراع

۳۸. از دلایل پرداختن به پروانه‌های ثبت اختراع یو. اس. (US patent)، اعتبار مرکز صادرکننده آن (داره ثبت اختراع و علائم تجاری امریکا) و حجم گسترده‌ای است که این منابع به خود اختصاص داده‌اند.

39. USPC (US patent classification)

40. Indented

۴۱. فناوری نانو یکی از فناوری‌های مورد توجه در عرصه ملی و بین‌المللی است. به همین دلیل، در پژوهش حاضر یکی از رده‌های فرعی مربوط به این فناوری (۹۷۷/۷۷۴) مورد بررسی قرار گرفته است.

42. <http://rapid-i.com/content/view/181/190/>

43. <http://cpan.uwinnipeg.ca/htdocs/Algorithm-FuzzyCmeans/Algorithm/FuzzyCmeans.html>

44. Fuzzy c-means clustering

45. Perl

46. [www.evalita.it/sites/evalita.fbkeu/files/doc2011/weps2007](http://www.evalita.it/sites/evalita.fbkeu/files/doc2011/weps2007)

-scorer\_1.1.tar.gz

47. WePS

48. Test collection

رده ۹۷/۷۴، عبارت ۹۷/۷۴/cd در جست‌وجوی پیشرفته پایگاه مربوط به پروانه‌های ثبت اختراع اداره یو. اس. پی. تی. آ. به نشانی: <http://patft.uspto.gov/netahtml/PTO/search-adv.htm> مورد جست‌جو قرار گرفت. از میان ۱۰۴ خوشه، تعداد ۱۸ خوشه (پروانه‌های مربوط به رده ۹۷/۷۴) به دلیل یکسان بودن با پروانه‌های مورد استناد و تعداد ۱۱ خوشه نیز به دلیل عدم استناد به پروانه ثبت اختراع یو. اس. پی. تی. آ. از خوشه‌بندی مرجع حذف شد. در نهایت تعداد ۷۵ پروانه ثبت اختراع (خوشه) به دست آمد. تعداد پروانه‌های ثبت اختراع که در این پروانه‌ها مورد استناد قرار گرفته است برابر با ۷۱۷ پروانه ثبت اختراع است که در این تحقیق به عنوان مجموعه آزمایشی مورد استفاده قرار گرفته است.

پس از مشخص شدن مجموعه آزمایشی، براساس شماره پروانه‌های ثبت اختراع (با استفاده از پایگاه‌های اطلاعاتی یو. اس. پی. تی. آ.، گوگل پتنتس<sup>۴۹</sup>، فری پتنتس آنالاین<sup>۵۰</sup>)، (حدود ۵۰۰۰۰ استناد) و درج آنها در قالب جدولی در نرم‌افزار اکسل اقدام به استخراج استنادها شده است. در نهایت جدولی تشکیل شد که قسمت‌هایی از آن در زیر قابل مشاهده است.

A	B	C	D	E	F	G	H	I	J	K
20	7431867	2932629	5489449	5490807	5741372	6342763	20050057867	20070054984	20070056463	
21	7422790	4207229	4225198	4251303	4304796	4322901	4326144	4440213	4467829	4617583
22	7421770	4476292	4769292	5347190	5834893	5955837	6091195	6777871	6787796	6920257
23	7405002	3656981	4751194	5549973	5731018	5781263	5788738	5879715	5906670	5965212
24	7402032	3995111	4211467	4329374	4557551	4688882	4891547	4818050	4854859	4894818
25	7399675	4184207	4751558	4785137	4813094	4860070	5196722	5252845	5315142	5411905
26	7399429	20020180195	20020187947	20050129947						
27	7394092	20020189108								
28	7390568	4882609	5021360	5081511	5262157	5293050	5354707	5395791	5412231	5422409
29	7382017	5043940	5434825	5747396	5977295	5959896	6054349	6090666	6139426	6259620
30	7378251	6114018	6306610	7138957	20010021078	200200403045				
31	7361821	3772200	4053433	4681195	4683202	4800159	5401511	5405706	5427930	5451505
32	7361543	4297995	6320794	6444545	6794712	6808992	6956265	6964902	7015090	7091089
33	7358525	3995111	4211467	4329374	4557551	4688882	4891547	4818050	4854859	4894818
34	7357018	5172930	5583443	5709940	6159742	6297063	6465782	6803840	6852679	6913790
35	7346251	20020191905	20030011309	20040113097	20060147169	20060151793				
36	7341514	5408115	5634584	5969183	6816414	6944902	20020190343	20050176202	20060189079	
37	7335345	4751200	5747180	5981353	6251303	6331110	6350515	6426513	6444543	6466808
38	7330963	4118485	4671084	5057513	5142185	5155051	5372675	5427948	5782908	6066204
39	7306823	3423101	3586541	3891519	3907566	3966568	4104134	4191794	4192721	4404422
40	7303937	5156316	5306345	6521948	6948442					
41	7274800	5390061	5465185	5549978	5640343	5650958	5729410	5757895	5764567	5843892
42	7259084	5141891	5181776	5238809	5308444	5438951	5471174	5879962	5959308	6207625
43	7256435	5277789	5772905	6256767	6294450	6432740	6504292	6891705	6897098	20020175408
44	7256416	20050064206								20040821842
45	7250651	5714766	6208000	6389343	6444545	6690019	20020078850			
46	7247929	41137623								
47	7238190	5212156	5254550	5879970	5891769	6030894	6091252	6129782	6371112	6537370
48	7234428	4642314	4944778	5034506	5151507	5187985	5305928	5558916	5596886	5602240
49	7208779	5128819	5783478	6417085	6412801	6790394	6743681	20020109142	20040178470	4642314
50	7201485	4992999	6328866	6328987	6413615	6447878	6461003	6508558	20020191291	

## جدول ۱

ستون A شماره پروانه ثبت اختراع و ستون‌های بعدی شماره پروانه‌ها و درخواست‌نامه‌های ثبت اختراع استناد شده

جدول ۱ مربوط به استنادهای موجود در هر پروانه ثبت اختراع است که در ستون اول (شماره پروانه ثبت اختراع) و در ستون‌های بعد «شماره پروانه‌ها و درخواست‌نامه‌های ثبت اختراع مورد استناد» درج شده است.

49. <http://www.google.com/patents>  
50. <http://www.freepatentsonline.com/search.html>

## ایجاد ماتریس استناد- پروانه ثبت اختراع

برای ایجاد ماتریس «استناد- پروانه ثبت اختراع» جدول ۱ به نرم‌افزار ریپدماینر وارد شده و وزن‌دهی استنادها براساس روش  $tf-idf$ <sup>۵۱</sup> انجام گرفت. این روش یکی از متداول‌ترین روش‌های مورد استفاده برای وزن‌دهی خصیصه‌هاست که بخش اول آن ( $tf$ ) از سوی سالتون و بخش دوم آن ( $idf$ ) از سوی اسپارک جونز پیشنهاد شده است (تیان، و تنگ<sup>۵۲</sup>، ۲۰۱۱).  $tf$  برابر با سهم حضور یک واژه در یک مدرک است و  $idf$  ضریبی است که با توجه به میزان رخداد هر خصیصه در کل مجموعه محاسبه می‌شود.  $idf$  برابر با لگاریتم معکوس میزان حضور واژه در کل مجموعه مدارک است و سبب می‌شود که به خصیصه‌های کم‌رخداد در کل مجموعه وزن بیشتری اختصاص یابد. در این روش وزن هر خصیصه از ضرب «بسامد حضور یک خصیصه در یک مدرک» ( $tf$ ) در «لگاریتم تعداد کلیه مدارک بخش بر تعداد مدارک حاوی هر خصیصه» ( $idf$ ) به دست می‌آید (تی اف- آی دی اف<sup>۵۳</sup>، ۲۰۱۱).

## نحوه خوشه‌بندی پروانه‌های ثبت اختراع

در تحقیق حاضر، برای بررسی تأثیر خصیصه اشتراک استنادی (زوج‌های کتابشناختی) از خوشه‌بندی سی- مینز (میانگین C) استفاده شده است. در این خوشه‌بندی میزان عضویت در یک خوشه مقادیر میان صفر و یک را می‌تواند به خود اختصاص دهد. برای ایجاد خوشه‌بندی فازی سی- مینز مراحل زیر دنبال می‌شود (ودینگ<sup>۵۴</sup>، ۲۰۰۹):

۱. انتخاب تعداد دلخواه  $k$  خوشه برای مجموعه‌ای شامل  $N$  مدرک<sup>۵۵</sup>، به طوری که  $k < N$ ؛

که در اینجا  $K$  برابر است با ۷۵ خوشه و  $N$  برابر با ۷۱۷ پروانه ثبت اختراع است؛

۲. ایجاد یک نقطه مرکزی آغازین به ازای هر یک از  $K$  خوشه؛

۳. محاسبه فاصله هر یک از  $N$  مدرک تا هر یک از  $K$  خوشه؛

۴. تعیین یک عضویت فازی یا کسری در هر یک از  $K$  خوشه برای هر یک از  $N$

مدرک؛

۵. یافتن نقطه مرکزی جدیدی برای هر یک از  $k$  خوشه از طریق یافتن میانگین وزنی

رکوردها<sup>۵۶</sup>؛

۶. تکرار مراحل ۳، ۴، و ۵ تا هنگامی که تغییری در عضویت خوشه‌ای به وجود نیاید (یا تا

هنگام رسیدن به معیار همگرایی).

## تعیین درجه عضویت مدارک در هر یک از خوشه‌ها

درجه عضویت فازی، از محاسبه فاصله‌های مدارک از مرکز خوشه‌ها به دست می‌آید. هنگامی

51. Term frequency-inverse

document frequency

52. Tian & Tong

53. tf-idf

54. Wedding

۵۵. در توصیف الگوریتم از واژه data استفاده

شده است، زیرا کاربرد خوشه‌بندی تنها به

خوشه‌بندی مدارک محدود نیست، اما از آنجا که

در اینجا خوشه‌بندی برای مدارک مورد استفاده

قرار گرفته است، واژه «مدرک» مورد استفاده

قرار گرفته است.

۵۶. Records: در اینجا منظور همان

پروانه‌های ثبت اختراع هستند که در صورتی که

بخواهیم میانگین وزنی از آن‌ها بگیریم باید از

تک تک مؤلفه‌هایشان میانگین وزنی بگیریم تا

بردار خوشه جدید به دست آید.



که مدرکی به مرکز یک خوشه نزدیک تر از مرکز دیگر خوشه‌ها باشد، درجه عضویت مدرک در آن خوشه بیشتر و هنگامی که مدرک دورتر باشد درجه عضویت کمتری خواهد داشت. مجموع مقادیر عضویت فازی [یک مدرک] نیز برابر با یک است (ودینگ، ۲۰۰۹).  
درجه عضویت از طریق فرمول زیر (ودینگ، ۲۰۰۹) محاسبه می‌شود:

$$u_k = \frac{1}{\sum_{i=1}^j \left(\frac{d_k}{d_i}\right)^p}$$

که در آن  $k$  یکی از  $j$  خوشه،  $d$  فاصله و  $p$  توان است. توان ( $p$ ) خود از فرمول زیر (ودینگ، ۲۰۰۹) به دست می‌آید:

$$p=2(m-1)$$

که در آن  $m$  یک توان<sup>۵۷</sup> فازی است به طوری که  $1 < m < \infty$  است. مقدار  $m$  بر درجه فازی بودن خوشه‌ها تأثیر گذار است. هنگامی که  $m \rightarrow 1$  (م نزدیک به ۱ باشد) درجه فازی شدن خوشه‌ها کمتر و هنگامی که  $m \rightarrow \infty$  (به سمت بی‌نهایت میل کند) خوشه‌ها به‌طور فزاینده‌ای فازی خواهند شد. برای مقادیر نزدیک به بی‌نهایت، مقدار عضویت‌های فازی یک مدرک در همه خوشه‌ها به برابری می‌گراید. معمولاً برای  $m$  مقدار ۲ در نظر گرفته می‌شود (ودینگ، ۲۰۰۹).

پس از به‌دست آوردن خصیصه‌های وزن داده شده در نرم‌افزار ریپیدماینر، از برنامه نوشته شده به زبان پریل برای ایجاد خوشه‌بندی فازی سی-مینز استفاده شده است. از آنجا که خوشه‌بندی مرجع دارای ۷۵ خوشه بود، در تنظیمات خوشه‌بندی، تعداد خوشه برابر با ۷۵ خوشه (طی ۱۰۰ تکرار) تعیین شد.

خروجی این برنامه، جدولی است که در آن میزان درجه عضویت هر پروانه ثبت اختراع را به هر یک از ۷۵ خوشه نشان می‌دهد.

جدول ۲ نشان دهنده عضویت‌های فازی پروانه‌های ثبت اختراع در خوشه‌بندی حاصل از زوج‌های کتابشناختی است. در ستون  $A$  شماره پروانه‌های ثبت اختراع قرار گرفته و در ستون‌های بعدی نام هر خوشه قرار دارد.

آمار مربوط به میزان درجه عضویت پروانه‌های ثبت اختراع در هر خوشه نیز در جدول ۳ آمده است.

57. Exponent

1	A	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
2	2155658***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
3	3203319***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
4	3406228***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
5	3410880***	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
6	3450673***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
7	3485806***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
8	3488327***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
9	3488389***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
10	3499032***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
11	3511894***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
12	3527802***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
13	3555095***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
14	3565868***	0	0	0	0	0	0	0	0	0	0
15	3567725***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
16	3583950***	0	0	0	1	0	0	0	0	0	0
17	3598790***	0.0109	0.0109	0.0109	0.195	0.0109	0.0109	0.0109	0.0109	0.0109	0.0109
18	3641145***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
19	3664990***	0	0	0	0	0	0	0	0	0	0
20	3671473***	0	0	0	0	0	0	0	0	0	0
21	3676453***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
22	3704128***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
23	3705992***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
24	3706797***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
25	3706800***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
26	3715343***	0	0	0	0	0	0	0	0	0	0
27	3719933***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
28	3728387***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133
29	3756994***	0.0132	0.0132	0.0132	0.0132	0.0132	0.0132	0.0132	0.0132	0.0132	0.0132
30	3772200***	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133	0.0133

### جدول ۲

بخشی از ماتریس مربوط به  
عضویت فازی هر مدرک  
در هر خوشه براساس  
زوج‌های کتابشناختی

### جدول ۳

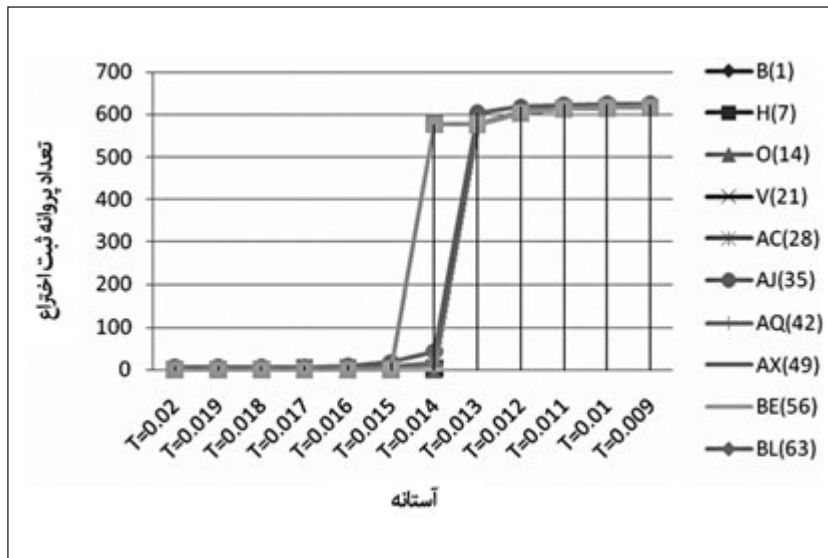
درجه عضویت در هر  
خوشه-زوج‌های کتابشناختی

	درجه عضویت در خوشه (اشتراک استنادی)
Min	۰
Max	۱
Average	۰/۰۱۳۳۱۴
Median	۰/۰۱۳۳
Mode	۰/۰۱۳۳
Std	۰/۰۴۰۵۱

## یافته‌ها

پس از مشخص شدن درجه‌های عضویت هر پروانه ثبت اختراع در هر ۷۵ خوشه، آستانه‌ای برای ارزیابی مانعیت محور و آستانه‌ای برای ارزیابی جامعیت محور تعیین شده که این آستانه با استفاده از مطالعه بر روی تعدادی از خوشه‌ها تعیین شده است. برای تعیین آستانه مناسب،

آستانه‌های موجود در بازه ۰/۰۲ تا ۰/۰۰۹ مورد بررسی قرار گرفته‌اند. این بازه با توجه به شاخص‌های مربوط به میانگین و میانه و نیز انحراف معیار جدول ۱ انتخاب شد. نمودارهای مربوط به تغییرات آستانه و تغییرات تعداد اعضای هر خوشه برای تعداد خوشه انتخابی در ادامه ارائه شده است.



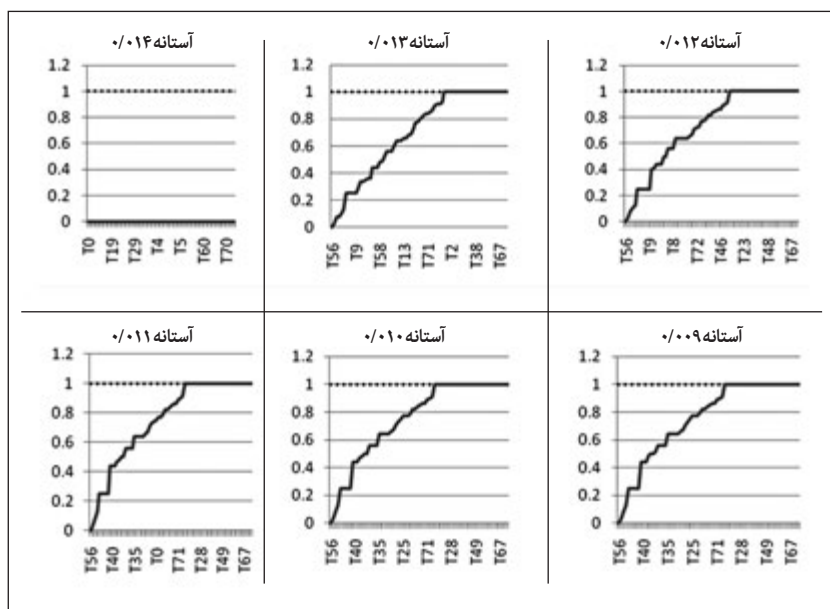
## نمودار ۱

زوج‌های کتابشناختی-تغییرات  
تعداد اعضای هر خوشه با توجه  
به تغییرات آستانه عضویت

طبق نمودار ۱، تغییرات تعداد اعضای اکثر خوشه‌ها از آستانه ۰/۰۱۵ به ۰/۰۱۳ شدت گرفته (شیب منحنی) و پس از آستانه ۰/۰۱ تغییر چندانی نکرده است. در پژوهش حاضر آستانه ۰/۰۱۴ (درجه عضویت ۰/۰۱۴ و بیشتر) برای ارزیابی مانعیت محور، و آستانه ۰/۰۰۹ (درجه عضویت ۰/۰۰۹ و بیشتر) برای ارزیابی جامعیت محور، مبنای قرار گرفته است، و برای بررسی روند تغییرات از مانعیت محوری به سمت جامعیت محوری، آستانه‌های ۰/۰۱۳، ۰/۰۱۲، ۰/۰۱۱، ۰/۰۱۰ نیز مورد بررسی قرار گرفته است. به‌ازای اعمال هر درجه‌ای از آستانه برای خوشه‌بندی به‌وجود آمده، میزان جامعیت و مانعیت گسترش یافته بی‌کیوبد محاسبه شد. نمودار ۲ نشان‌دهنده میزان جامعیت و مانعیت گسترش یافته بی‌کیوبد در آستانه‌های گوناگون است.

## مُودار ۲

میزان جامعیت و مانعیت  
گسترش یافته‌بی کیوبد  
در آستانه‌های گوناگون



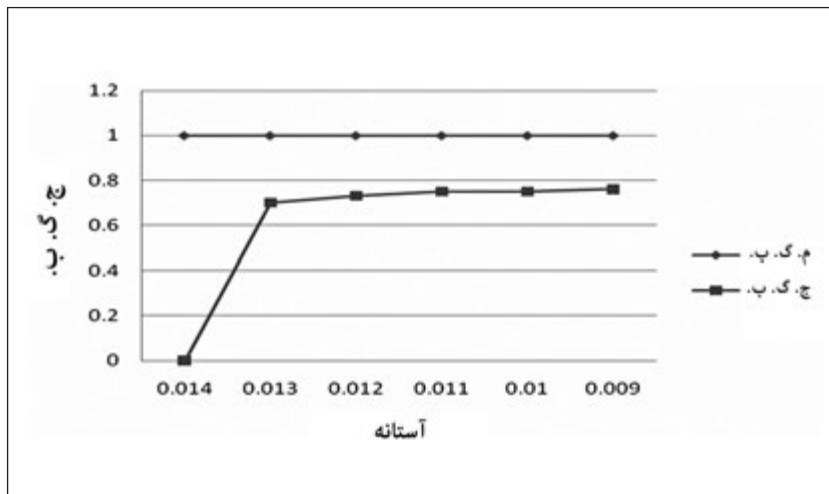
## جدول ۴

مقادیر نهایی جامعیت و مانعیت  
گسترش یافته‌بی کیوبد در  
آستانه‌های اعمال شده

آستانه	م. گ. ب.	ج. گ. ب.
۰/۰۱۴	۱	۰
۰/۰۱۳	۱	۰/۷
۰/۰۱۲	۱	۰/۷۳
۰/۰۱۱	۱	۰/۷۵
۰/۰۱۰	۱	۰/۷۵
۰/۰۰۹	۱	۰/۷۶
میانگین	۱	۰/۶۱۵
میانه	۱	۰/۷۴

همان‌طور که جدول ۴ نشان می‌دهد، در آستانه ۰/۰۱۴ مقدار جامعیت گسترش یافته صفر است، اما مانعیت آن عدد ۱ را نمایش می‌دهد. همچنین به ترتیب هر چه از آستانه ۰/۰۱۴ به سمت پایین‌تر سیر می‌شود، یعنی در آستانه‌های ۰/۰۱۳، ۰/۰۱۲، ۰/۰۱۱، ۰/۰۱۰ و ۰/۰۰۹، مقدار جامعیت گسترش یافته بی کیوبد افزایش می‌یابد.

در آستانه‌های تعیین شده، میانگین میزان مانعیت گسترش یافته بی‌کیوبد برابر با ۱ و میانگین میزان جامعیت آن برابر با ۰/۶۱۵ است. طبق جدول ۴، هرچه از آستانه تعریف شده کاسته می‌شود، مقدار جامعیت گسترش یافته افزایش می‌یابد، اما نرخ رشد و تغییرات آن اندک است. نمودار ۳ نشان‌دهنده تغییرات جامعیت و مانعیت گسترش یافته بی‌کیوبد برحسب تغییر در آستانه و حرکت از جامعیت محوری در خوشه‌بندی به سمت مانعیت محوری است.



### نمودار ۳

تغییرات جامعیت و مانعیت گسترش یافته برحسب تغییرات آستانه

باتوجه به نمودار ۳، هرچه از مانعیت محوری (آستانه ۰/۰۱۴) در خوشه‌بندی به سمت جامعیت محوری (آستانه ۰/۰۰۹) پیش می‌رویم، میزان جامعیت گسترش یافته افزایش می‌یابد. با این حال، در تمامی آستانه‌های تعیین شده، میزان مانعیت گسترش یافته مقداری ثابت و برابر با ۱ را به خود اختصاص داده است.

همچنین نمودار ۳ حاکی از آن است که خوشه‌بندی با استفاده از زوج‌های کتابشناختی، در طیف گسترده‌تری از آستانه قابل تعریف است. این بدان معنی است که ساختار خوشه‌ای در خوشه‌بندی با استفاده از زوج‌های کتابشناختی از انسجام و عملکرد مطلوبی برخوردار است.

## نتیجه‌گیری

جنبه حقوقی در ثبت اختراعات سبب شده است که نیازهای اطلاعاتی در جست‌وجوی پروانه‌های ثبت اختراع نیازمند جامعیت بیشتر در نتایج جست‌وجو باشد تا افرادی که در جهت بررسی قابلیت ثبت اثری به‌عنوان اختراع هستند بتوانند تمامی منابع مرتبط را مورد

بررسی قرار دهند. یافته‌های پژوهش حاضر حاکی از آن بود که خوشه‌بندی با استفاده از زوج‌های کتابشناختی، عملکردی مطلوب (بیش از ۵۰ درصد) داشته و این مطلوبیت در طیف گسترده‌ای از آستانه‌های تعیین شده (میزان جامعیت محوری) برقرار است. بنابراین، به نظر می‌رسد استفاده از این خصیصه در بازیابی اطلاعات پروانه‌های ثبت اختراع نتایج مطلوبی در پی داشته باشد.

عملکرد مطلوب خصیصه استناد در پژوهش فوجی (۲۰۰۷) نیز نشان داده شده است. با این حال بازیابی تعداد زیادی از منابع نامرتبط در کنار منابع مرتبط هنگام استفاده از استناد (تیوانا و هورویتز، ۲۰۰۹) را نمی‌توان تأییدی بر مطلوبیت استفاده از این خصیصه دانست. یافته‌های پژوهش حاضر همچنین نشان داد که علاوه بر رده‌بندی خودکار (لای و وو، ۲۰۰۵؛ لی و دیگران، ۲۰۰۷) و بازیابی اطلاعات (فوجی، ۲۰۰۷)، استفاده از استناد نیز در خوشه‌بندی پروانه‌های ثبت اختراع نقشی کارآمد دارد. بنابراین شاید بتواند روشی مؤثر در ترسیم نقشه‌های فناوری با استفاده از خوشه‌بندی و نیز بازیابی‌های خوشه‌محور باشد. البته باید به این مسئله نیز اشاره کرد که با گذشت زمان حجم استنادها افزوده خواهد شد. به‌طور مثال، در مورد پروانه‌های ثبت اختراع، هر روزه تعدادی اختراع به ثبت می‌رسد و تعدادی از منابع مورد استناد قرار می‌گیرند. این امر سبب افزایش حجم استنادها و در نتیجه خصیصه‌های مورد استفاده در خوشه‌بندی پروانه‌های ثبت اختراع خواهد شد. این امر، پژوهش در باب کاهش حجم مورد نیاز برای نگهداری خصیصه‌ها و شکل‌های مختلف پردازش آنها را ضروری می‌سازد. بنابراین، پیشنهاد می‌شود پژوهش‌هایی در باب کاهش حجم خصیصه‌های استنادی انجام شود تا بهترین راه که با جامعیت بالاتری همراه باشد مورد استفاده قرار گیرد.

شایان ذکر است که پژوهش حاضر به بررسی پروانه‌های ثبت اختراع مربوط به رده نانو تکنولوژی و استفاده از الگوریتم فازی پرداخته است و برای اینکه بتوان نتایج آن را به دیگر حوزه‌ها نیز تعمیم داد، شاید لازم باشد تحقیقات مشابه دیگری نیز در سایر حوزه‌ها انجام شود.

## منبع

Bashir, S.; Rauber, A. (2009). "Improving retrievability of patents with cluster-based pseudo-relevance feedback documents selection". In: *CIKM '09: Proceeding of the 18th ACM conference on Information and knowledge management*, (pp. 1863–1866), New York, NY, USA. ACM. Retrieved February 19, 2011 from: [http://publik.tuwien.ac.at/files/PubDat\\_182418.pdf](http://publik.tuwien.ac.at/files/PubDat_182418.pdf)

- Bonino, D.; Ciaramella, A.; Corno, F. (2010). "Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics". *World Patent Information*, 32(1):30 – 38.
- Dhillon I.; Kogan J.; Nicholas CH. (2003). "Feature selection and document clustering". In: M. Berry (Ed.), *A Comprehensive Survey of Text Mining* (pp. 73-100). Springer-Verlag. Retrieved July 29, 2011, from: <http://www.cs.umbc.edu/csee/research/cadip/2002Symposium/kogan.pdf>
- Fujii, A. (2007). "Integrating content and citation information for the NTCIR-6 patent retrieval task". In: *Proceedings of NTCIR-6 Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access* (pp.377-380). Tokyo, Japan.
- Graf, E.; Azzopardi, L. (2008). "A methodology for building a patent test collection for prior art search". In: *Proceedings of the 2nd International Workshop on Evaluating Information Access (EVIA), December 16, 2008, Tokyo, Japan* (pp. 60–71). Retrieved January 27, 2010 from: <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings7/pdf/EVIA2008/11-EVIA2008-GrafE.pdf>
- Hideo, J. ;Azzopardi, L.; Vanderbauwhede, W. (2010)."A survey of patent users". In: *Proceeding of the third symposium on Information interaction in context, August 18-22*, (pp. 13-22). New Brunswick, New Jersey, USA, ACM doi>10.1145/1840784.1840789
- Kang, I.; et al. (2007). "Cluster-based patent retrieval". *Information Processing and Management*, 43 (5): 1173-1182.
- Lai ,K.K. ; Wu, SH.J. (2005)."Using the patent co-citation approach to establish a new patent classification system". *Information Processing and Management* , 41: 313–330.
- Li, X.; et al. (2007). "Automatic patent classification using citation network information: An experimental study in nanotechnology". In: *JCDE'07: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, (pp. 419–427): New York, NY, USA. ACM. doi>10.1145/1255175.1255262
- Tan, P.N.; Steinbach, M.; Kumar, V. (2006). "Cluster analysis: Basic concept and algorithm". In: P.N.Tan, M. Steinbach and V. Kumar. *Introduction to Data Mining*, (pp.487-568). Boston, MA: Addison-Wesley Longman Publishing.
- tf-idf (2011, August 30). Retrieved June (?) 01(?), 2011, from Wikipedia, the free encyclopedia: <http://en.wikipedia.org/wiki/Tf%E2%80%93idf>

- Tian, X.; Tong, W. (2011). "An improvement to TF-IDF: Term distribution based term weight algorithm". *Journal of Software*, 6(3): 413-420.
- Tiwana, S.; Horowitz, E. (2009). "FindCite: Automatically finding prior art patents". In: *Proceedings of the 2nd international workshop on Patent information retrieval*, pp. 37-40. Retrieved March 03, 2011 from: [http://delivery.acm.org/10.1145/1660000/1651352/p37-tiwana.pdf?ip=217.218.83.9&CFID=39566988&CFTOKEN=46635715&\\_\\_acm\\_\\_=1315641819\\_3628c1ae1982ae3f1f9b5747a3f756](http://delivery.acm.org/10.1145/1660000/1651352/p37-tiwana.pdf?ip=217.218.83.9&CFID=39566988&CFTOKEN=46635715&__acm__=1315641819_3628c1ae1982ae3f1f9b5747a3f756)
- United States Patent and Trademark Office (USPTO) (2010). "Manual of Patent Examining Procedure (MPEP)". (eighth edition). Retrieved January 09, 2011 from: [http://www.uspto.gov/web/offices/pac/mpep/pdf\\_download\\_search\\_instructions.doc](http://www.uspto.gov/web/offices/pac/mpep/pdf_download_search_instructions.doc)
- Wanagiri, M. Z.; Adriani, M. (2010). "Prior art retrieval using various patent document fields contents". In: CLEF-2010(Notebook Papers/LABs/Workshops). Retrieved September 03, 2011 from: [http://www.clef2010.org/resources/proceedings/clef2010labs\\_submission\\_85.pdf](http://www.clef2010.org/resources/proceedings/clef2010labs_submission_85.pdf)
- Wedding, D. K. (2009). "Extending the data mining software packages SAS enterprise miner and SPSS clementine to handle Fuzzy cluster membership: Implementation with Examples". Master of Science Thesis in Data Mining, Central Connecticut State University. Retrieved May 13, 2011 from: <http://web.ccsu.edu/datamining/Data%20Mining%20Theses/Don%20Wedding%20thesis.pdf>
- Xue, X.; Croft W. B. (2009). "Transforming patents into prior-art queries". In: *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, (pp. 808-809): Boston, MA, USA. ACM. [doi>10.1145/1571941.1572139]

استناد به این مقاله: کرمانی، آناهیتا؛ نشاط نرگس (۱۳۹۱). «ارزیابی زوج‌های کتابشناختی در خوشه‌بندی پروانه‌های ثبت اختراع». *فصلنامه مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۲۳(۳).