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 گسترش‌یافته 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 برمبنای زوج‌های کتابشناختی، استفاده از زوج‌های کتابشناختی در فرآیند سازماندهی و بازیابی پروانه‌های ثبت اختراع نیز می‌تواند عملکرد مطلوب داشته باشد.

کلیدواژه‌ها

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

ارزیابی زوج‌های کتابشناختی در خوشه‌بندی پروانه‌های ثبت اختراع

آناهیتا کرمانی[1] | نرگس نشاط[2]

دریافت: ۱۳۹۰/۸/۴ 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» ازجمله 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 غیربدیهی بودن یک اختراع یا رد اعتبار آن، مورد جست‌وجو قرار گیرند. ارزیاب[3] مأمور به بررسی ماهوی یک اختراع به دلیل بررسی تازگی و غیربدیهی بودن آن، و شخصی که مدعی تجاوز به حق اختراع خود است باید برای رد اعتبار اختراع ثبت شده، به جست‌وجو در منابع و ازجمله پروانه‌های ثبت اختراع بپردازد (گراف، و آزوپاردی[4]، ۲۰۰۸؛ اداره ثبت اختراع و علائم تجاری امریکا[5]، ۲۰۱۰؛ بونینو، شیاراملا، کورنو[6]، ۲۰۱۰، نقل در هیدئو، آزوپاردی، وندرباووهد[7]، 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, & Corno
7. Hideo, Azzopardi, & Vanderbauwhede

بازیابی مناسب پروانه‌های ثبت اختراع را نشان می‌دهد. در این راستا، پژوهش‌های متفاوتی پیرامون سازماندهی و بازیابـی اطلاعـات پروانه‌های ثبت اختـراع انجـام شده که در برخی از آنها (کانگ[8] و دیگران، ۲۰۰۷؛ بشیر، و رابر[9]، ۲۰۰۹) نیز به روش خوشه‌بندی پرداخته شده است.

خوشه‌بندی روشـی است که برای گروه‌بنـدی موجودیت‌های[10] (مدارک) مشابه مورد استفاده قرار می‌گیرد. در این روش، مدارک در گروه‌هایی از پیش تعیین نشده به نام خوشه قرار می‌گیرند؛ به‌طوری که مدارک مشابه در کنار یکدیگر و مدارک نامشابه دور از یکدیگر قرار گیرند (تن، اشتینباخ، و کومار[11]، ۲۰۰۶).

عوامل گوناگونی در خوشه‌بندی مؤثر است؛ یکی از این عوامل مربوط به نوع خصیصه‌ای[12] 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[13]، ۲۰۰۳). به این دلیل، انتخاب درست خصیصه برای مدارک در نتایج خوشه‌بندی تأثیرگذار خواهد بود.

غالباً خصیصه‌های گوناگونی برای بازنمایی[14] 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[15]. از این رو، می‌توانند به‌عنوان 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[16] (۲۰۰۵)، لی[17] و دیگران (۲۰۰۷)، فوجی[18] (۲۰۰۷) اشاره کرد که عنصر استناد را به‌نوعی 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»[19] استفاده از رویکرد هم‌استنادی را به‌منظور رده‌بندی پروانه‌های ثبت اختراع پیشنهاد کرده‌اند. نتایج ارزیابی این مدل با استفاده از روش تجربی در مجموعه‌ای آزمایشی شامل پروانه‌های مربوط به نانوتکنولوژی، حاکی از مفید بودن اطلاعات استنادی در سازماندهی پروانه‌های ثبت اختراع بوده است.

کانگ و دیگران (۲۰۰۷) در مقالـه‌ای با عنوان «بازیابـی خوشـه- محور برای پروانه‌های ثبت اختراع»[20] به مطالعـه 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
۱۵. البته باید توجه داشت که نظام های ثبت اختراع در مراکز گوناگون متفاوت است و توصیف یاد شده از استناد در مراکزی صدق می کند 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»[۲۱]، 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»[۲۲] 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 درنظرگرفته می‌شود؛ و حالتی که نمره استنادی در N مدرک برتر حاصل از بازیابی مبتنی بر متن محاسبه می‌شود. در این مدل، به‌منظور نمره‌دهی ثبت اختراعات از نظر اطلاعات استنادی از روش پیج رنک [۲۳] استفاده شده است. نتایج این پژوهش تجربی نشان داده که تلفیق اطلاعات استنادی در حالت اول (محاسبه نمره استنادی بدون درنظرگرفتن N مدرک برتر) اثربخش‌تر از حالت دوم عمل کرده، اما به‌طورکلی، تلفیق اطلاعات استنادی و متنی در بازیابی اثربخش بوده است.

تیوانا، و هورویتز[۲۴] (۲۰۰۹) در «فایندسایت – یافتن خودکار پروانه‌های ثبت اختراع پیشین»[۲۵] 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 وجو 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[۲۶](۲۰۰۹) در مقاله خود با عنوان «تبدیل پروانه‌های ثبت اختراع به عبارت‌های جستجو برای آثار پیشین»[۲۷] 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 درنظرگرفته شده است. نتایج حاصل از تحقیق تجربی بر روی مجموعه حاوی پروانه‌های ثبت اختراع اداره یو. اس. پی. تی. اُ.[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[۲۹] 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.

ونگیری و آدریانی[۳۰] (۲۰۱۰) در پژوهش خود با عنوان «بازیابی آثار پیشین با استفاده از محتوای نواحی مختلف در مدارک پروانه ثبت اختراع»[۳۱]، عملکرد بازیابی به‌منظور رد اعتبار[۳۲] 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)[۳۳] و NDCG است. همچنین میزان جامعیت در هر دو مجموعه موضوعات عام[۳۴] و خاص[۳۵] 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»[۳۶]، و روش دوم که «مدارک استنادکننده» به‌عنوان خصیصه‌ای برای مدارک «استناد شونده» محسوب شود «هم‌استنادی»[۳۷] 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.[۳۸] مورد استناد در پروانه‌های ثبت اختراع یو. اس. متعلق به رده ۹۷۷/۷۷٤ از رده‌بندی یو. اس. پی. سی.[۳۹] است. رده ۹۷۷ رده‌ای اصلی و مربوط به موضوع نانوتکنولوژی (فناوری نانو)؛ رده ۹۷۷/۷۷٤ رده‌ای فرعی و زیرمجموعه‌ای[٤۰] از رده فرعی ۹۷۷/۷۷۳؛ رده ۹۷۷/۷۷۳ نیز زیرمجموعه‌ای از رده ۹۷۷/۷۰۰ است. رده فرعی ۹۷۷/۷۰۰ مربوط به نانوساختارها و رده فرعی ۹۷۷/۷۷۳ شامل نانوذرات است[٤۱].

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

نرم‌افزار کد منبع باز رپیدماینر[٤۲] در وزن‌دهی به استنادها؛ برنامه[٤۳] خوشه‌بندی فازی[٤٤] نوشته‌شده به زبان برنامه‌نویسی پرل[٤۵]؛ و برنامه[٤٦] محاسباتی جامعیت و مانعیت گسترش‌یافته بی کیوبد که در گارگاه وبس[٤۷] معرفی و ارائه شده است.

در اجرای پژوهش، فرآیندهایی از جمله ایجاد مجموعه مورد آزمایش[٤۸]، استخراج استنادها، 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 استنادها

ارزیابی خوشه‌بندی معمولاً با استفاده از مجموعه‌های آزمایشی انجام می شود. 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 استنادهای 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 استنادهای موجود در آن به‌عنوان مدارک مرتبط با آن موضوع درنظرگرفته 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 ۱۰٤ پروانهٔ ثبت اختراع مربوط به رده ۹۷۷/۷۷٤ 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 (اداره 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 گرفته‌است.

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.fbk.eu/files/doc2011/weps2007_scorer_1.1.tar.gz

47. WePS

48. Test collection

رده ۹۷۷/۷۷٤، عبارت ccl/۹۷۷/۷۷٤ در جست‌وجوی پیشرفته پایگاه مربوط به پروانه‌های ثبت اختراع اداره یو. اس. پی. تی. اَ. به نشانی: http://patft.uspto.gov/netahtml/PTO/search-adv.htm مورد جست‌جو قرار گرفت. از میان ۱۰٤ خوشه، تعداد ۱۸ خوشه (پروانه‌های مربوط به رده ۹۷۷/۷۷٤) به دلیل یکسان بودن با پروانه‌های مورد استناد و تعداد ۱۱ خوشه نیز به دلیل عدم استناد به پروانه ثبت اختراع یو. اس. از خوشه‌بندی مرجع حذف شد. درنهایت تعداد ۷۵ پروانه ثبت اختراع (خوشه) به‌دست 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. او.، گوگل پتنتس[49]، فری‌پتنتس آنلاین[50])، (حدود ۵۰۰۰۰ استناد) و درج آنها در قالب جداولی در نرم‌افزار اکسل اقدام به استخراج استنادها شده است. درنهایت 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	5690807	5741372	6582763	20050057867	20070054984	20070056465		
21	7422790	6207229	6225198	6251303	6306736	6322901	6326144	6440213	6607829	6617583	20030136943
22	7417370	4476292	4769292	5247190	5834893	5955837	6091195	6777871	6787796	6920257	6965197
23	7405002	3656981	4751194	5549973	5751018	5783263	5788738	5879715	5906670	5965212	5990479
24	7402832	3995311	4211467	4332974	4557551	4688882	4693547	4818050	4856859	4894818	4906064
25	7399675	4184207	4751558	4785337	4833094	4860070	5196722	5252845	5315142	5411905	5432365
26	7399429	20020160195	20020187347	20050129947							
27	7394092	20020199108									
28	7390568	4882609	5021360	5081511	5262357	5293050	5354707	5395791	5412231	5422489	5434878
29	7382017	5043940	5434825	5714766	5937295	5959896	6054349	6090666	6139626	6159620	6207229
30	7378151	6114038	6306610	7318957	20010023078	20020045045					
31	7361821	3772200	4053433	4683195	4683202	4800159	5401511	5405766	5427930	5451505	5472881
32	7361543	6297095	6320784	6444545	6794712	6808992	6958265	6964902	7015090	7091089	7091130
33	7358525	3995311	4211467	4332974	4557551	4688882	4693547	4818050	4856859	4894818	4906064
34	7357018	5372930	5583443	5730940	6159742	6297063	6465782	6803840	6852673	6919730	6952952
35	7346251	20020191905	20030013109	20040131097	20060147169	20060151793					
36	7341914	5408115	5824584	5969383	6816414	6964902	20020190343	20050176202	20060189079		
37	7335345	4751200	5747180	5985353	6251303	6333110	6350515	6426513	6444143	6468808	6576291
38	7306963	4118485	4673584	5057313	5142385	5155051	5372675	5427648	5782908	6066204	6080928
39	7306823	3423301	3586541	3891519	3907566	3966568	4104134	4191794	4192721	4404422	4410758
40	7303937	5156336	5306345	6521048	6984842						
41	7274080	5390061	5465185	5549978	5640343	5650958	5729410	5757695	5764567	5841692	5862022
42	7259084	5141893	5183776	5238869	5308444	5438951	5473174	5879962	5959308	6107635	6291321
43	7256435	5277749	5772905	6256767	6294450	6432740	6504292	6893705	6897098	20020175408	20040023162
44	7256416	20050064206									
45	7250653	5714766	6208000	6288943	6444545	6690059	20020076850				
46	7249859	6137623									
47	7238595	5221556	5256550	5879970	5891769	6030894	6093252	6319782	6373112	6537370	6562736
48	7214428	4642334	4946778	5034506	5151507	5187085	5505928	5558991	5596086	5602240	5637258
49	7208793	5328859	5783478	6417085	6432801	6706594	6743681	20020109142	20040178470	4642334	4946778
50	7201485	4992699	6318866	6318987	6413615	6447878	6461003	6508558	20020191292		

جدول ۱

ستون A شماره پروانه ثبت اختراع و ستون‌های بعدی شماره پروانه‌ها و درخواست‌نامه‌های ثبت اختراع استنادشده

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- پروانهٔ ثبت اختراع» جدول ۱ به نرم‌افزار رپیدماینر وارد شده و وزن‌دهی استنادها براساس روش tf-idf[51] 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[52]، 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) به‌دست می‌آید (تی‌اف- آی‌دی‌اف[53]، 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[54]، ۲۰۰۹):

۱. انتخاب تعداد دلخواه k خوشه برای مجموعه‌ای شامل N مدرک[55]، به‌طوری 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؛

٤. 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[56]؛

٦. تکرار مراحل ۳، ٤، و ۵ تا هنگامی که تغییری در عضویت خوشه‌ای به‌وجود 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 به‌دست 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 درصورتی که بخواهیم میانگین وزنی از آن‌ها بگیریم باید از تک‌تک مؤلفه‌هایشان میانگین وزنی بگیریم تا بردار خوشه جدید به‌دست 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:

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

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[۵۷] فازی است به طوری که $1<m<\infty$ است. مقدار m بر درجه فازی بودن خوشه‌ها تأثیرگذار است. هنگامی که $m \to 1$ (m نزدیک به ۱ باشد) درجه فازی شدن خوشه‌ها کمتر و هنگامی که $m \to \infty$ (m 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 درنظرگرفته می‌شود (ودینگ، ۲۰۰۹).

پس از به‌دست آورن خصیصه‌های وزن داده شده در نرم‌افزار رپیدماینر، از برنامه نوشته شده به زبان پرل 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 جدول۳ آمده است.

57. Exponent

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

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

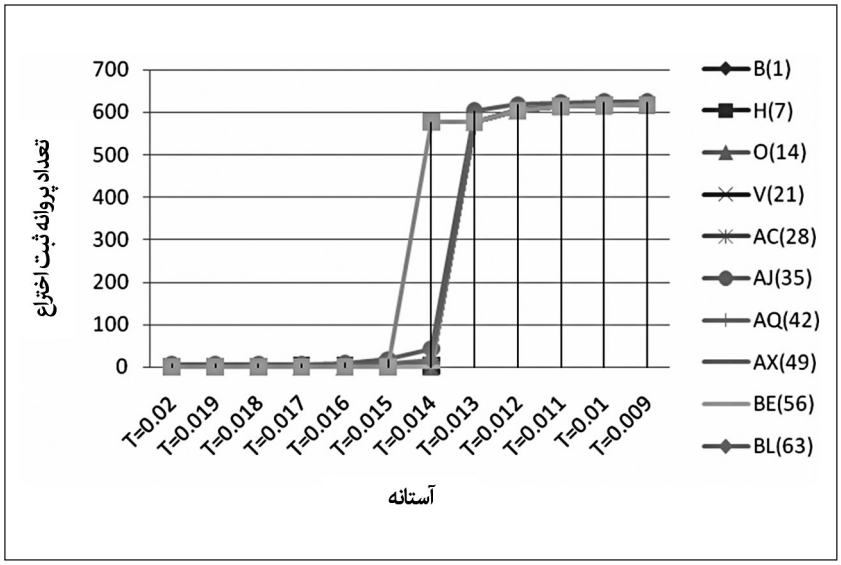
جدول ۳

درجه عضویت در هر خوشه-زوج‌های کتابشناختی

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 باتوجه 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.



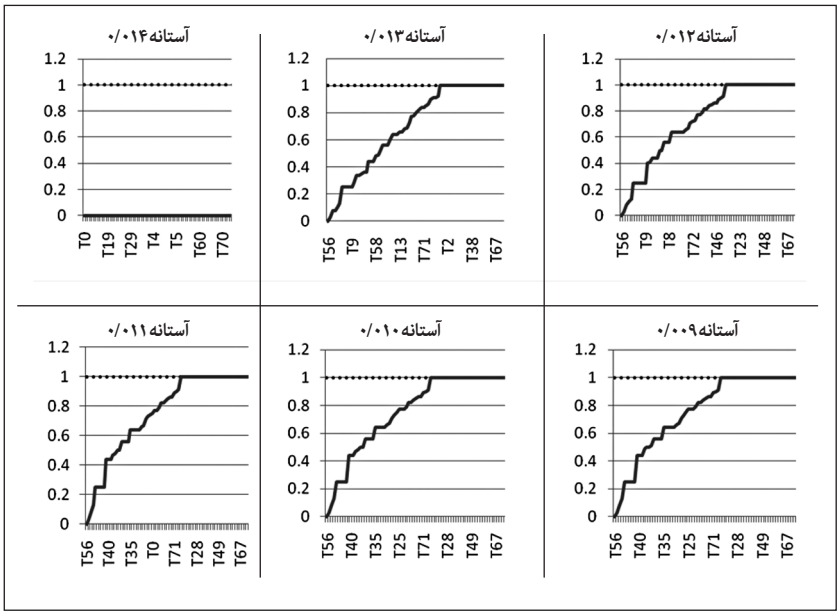
نمودار ۱

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 ۰/۰۱٤ (درجه عضویت ۰/۰۱٤ و بیشتر) برای ارزیابی مانعیت‌محور، و آستانه ۰/۰۰۹ (درجه عضویت ۰/۰۰۹ و بیشتر) برای ارزیابی جامعیت‌محور، مبنا 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 خوشه‌بندیِ به‌وجود آمده، میزان جامعیت و مانعیت گسترش یافتهٔ بی‌کیوبد محاسبه شد. نمودار۲ نشان‌دهنده میزان جامعیت و مانعیت گسترش‌یافته بی‌کیوبد در آستانه‌های گوناگون است.



نمودار ۲

میزان جامعیت و مانعیت گسترش‌یافته بی‌کیوبد در آستانه‌های گوناگون

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

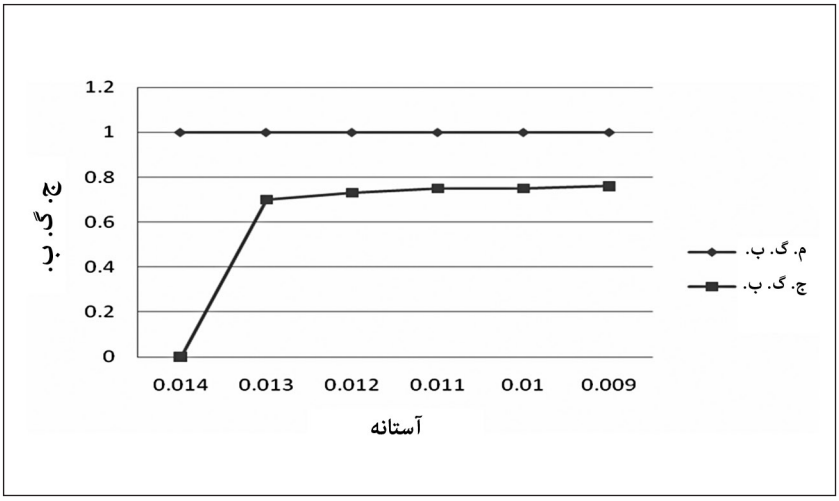
جدول ۴

مقادیر نهایی جامعیت و مانعیت گسترش‌یافته بی‌کیوبد در آستانه‌های اعمال شده

همان‌طور که جدول ٤ نشان می‌دهد، در آستانه ۰/۰۱٤ مقدار جامعیت گسترش‌یافته صفر است، اما مانعیت آن عدد ۱ را نمایش می‌دهد. همچنین به‌ترتیب هرچه از آستانه ۰/۰۱٤ به سمت پایین‌تر سیر می‌شود، یعنی در آستانه‌های ۰/۰۱۳، ۰/۰۱۲، ۰/۰۱۱، ۰/۰۱۰و ۰/۰۰۹ ، مقدار جامعیت گسترش‌یافته بی‌کیوبد افزایش می‌یابد.

در آستانه‌های تعیین شده، میانگین میزان مانعیت گسترش‌یافته بی‌کیوبد برابر با ۱ و میانگین میزان جامعیت آن برابر با ۰/٦١٥ است.

طبق جدول ٤، هرچه از آستانه تعریف شده کاسته می‌شود، مقدار جامعیت گسترش‌یافته افزایش می‌یابد، اما نرخ رشد و تغییرات آن اندک است. نمودار۳ نشان‌دهنده تغییرات جامعیت و مانعیت گسترش‌یافت بی‌کیوبد برحسب تغییر در آستانه و حرکت از جامعیت محوری در خوشه‌بندی به سمت مانعیت محوری است.



نمودار ۳

تغییرات جامعیت و مانعیت گسترش‌یافته برحسب تغییرات آستانه

باتوجه به نمودار ۳، هرچه از مانعیت محوری (آستانهٔ ۰/۰۱٤) در خوشه‌بندی به سمت جامعیت محوری (آستانهٔ ۰/۰۰۹) پیش می‌رویم، میزان جامعیت گسترش‌یافته افزایش می‌یابد. با این حال، در تمامی آستانه‌های تعیین شده، میزان مانعیت گسترش‌یافته 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، به‌نظر می‌رسد استفاده از این خصیصه در بازیابی اطلاعات پروانه‌های ثبت اختراع نتایج مطلوبی درپی 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 درنتیجه 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 نانوتکنولوژی 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.

منابع

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

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. Retreived 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 onEvaluating 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: ***JCDL'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_3628c1ae1982ae3f1f1f9b5747a3f756

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

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

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).