

مروری بر کارکردهای تحلیل هم‌واژگانی

حمید احمدی | فریده عصاره

چکیده

هدف: بررسی روش تحلیل هم‌واژگانی، یکی از این روش‌های حوزه علم‌سنجی.

یافته‌ها: در این پژوهش چارچوب‌ها، کارکردها، روش‌ها، و پیش‌فرض‌های موجود معرفی و الگوریتم‌ها و فناوری‌های مورد نیاز آن برای شناخت و استفاده از این روش در تحلیل محتوا بیان شد.

نتیجه‌گیری: تحلیل هم‌واژگانی با خلاصه‌سازی مدارک در واژه‌هایی قدرتمند و محاسبه رخداد و هم‌رخدادی، تشخیص دقیق‌تری نسبت به حوزه موضوعی ارائه می‌دهد و توان آن را دارد تا به کشف روابط پنهان میان مفاهیم و ایجاد سلسله‌مراتب در هستی‌شناسی‌ها بپردازد.

کلیدواژه‌ها

تحلیل هم‌واژگانی، تحلیل هم‌رخدادی واژگان، تحلیل محتوا، شبکه مفهومی، روش هم‌واژگانی

مروری بر کارکردهای تحلیل هم‌واژگانی

حمید احمدی^۱

فریده عصاره^۲

تاریخ دریافت: ۹۴/۱۰/۱۵

تاریخ پذیرش: ۹۴/۱۲/۱۲

مقدمه

سنجش و ارزیابی حوزه‌های علمی بدون استفاده از شاخص‌های کمی میسر نیست. این شاخص‌ها در روش‌های معمول تجزیه و تحلیل داده‌ها شامل زوج کتاب‌شناختی، تحلیل استنادی یا هم‌استنادی، تحلیل هم‌نویسندگی، و تحلیل هم‌واژگانی در حوزه کتاب‌سنجی و علم‌سنجی شناخته شده‌اند.

در پاسخ به پیشرفت‌های صورت گرفته در زمینه نگاشت هم‌استنادی توسط اسمال و گریفیث^۳ در سال ۱۹۷۴ افرادی مانند کالون، کورتیال، ترنر، و باین^۴ طرح نگاشت هم‌رخدادی واژگان را به عنوان زمینه پژوهش‌های دیگری برای مطالعه روابط معنایی در ادبیات علم و فناوری مطرح کردند (عصاره، سهیلی، و منصور، ۱۳۹۴، نقل از لیدسدورف^۵، ۱۹۸۹).

جدول ۱ متداول‌ترین طبقه‌بندی در روش‌های کتاب‌سنجی و علم‌سنجی طبق واحدهای تحلیلی مورد استفاده و روابط ایجادشده در میان آنهاست (کوبو^۶ و همکاران، ۲۰۱۱).

۱. دانشجوی دکتری علم اطلاعات و دانش‌شناسی، دانشگاه شهید چمران اهواز؛ عضو هیأت علمی گروه علم اطلاعات و دانش‌شناسی، دانشگاه رازی (نویسنده مسئول)

hamid_ahmadi@razi.ac.ir

۲. استاد گروه علم اطلاعات و دانش‌شناسی؛ دانشگاه شهید چمران اهواز

osareh.f@gmail.com

3. Small & Griffith

4. Callon, Courtial, Turner, & Bauin

5. Leydesdorff

6. Cobo

جدول ۱. طبقه‌بندی روش‌های کتاب‌سنجی و علم‌سنجی

روش‌ها	واحد تحلیل مورد استفاده	نوع رابطه
زوج کتاب‌شناختی	آثار مهم نویسنده، مدرک، مجلات مهم	مراجع مشترک میان آثار مهم نویسنده، مراجع مشترک میان مدارک، مراجع مشترک میان مجلات مهم

نوع رابطه	واحد تحلیل مورد استفاده	روش‌ها
هم‌رخدادی نویسندگان، هم‌رخدادی کشورها، هم‌رخدادی مؤسسات	نام نویسنده، نام کشور از وابستگی سازمانی، نام مؤسسه از وابستگی سازمانی	هم‌نویسندگی
نویسنده مورد استناد، مدارک مورد استناد، مجله مورد استناد	مأخذ نویسنده، مأخذ مجله	هم‌استنادی
هم‌رخدادی اصطلاحات یا واژه‌ها	کلمات کلیدی، یا اصطلاحات استخراج‌شده از عنوان، چکیده، یا مجموعه مدرک	هم‌واژگانی

کوبو و همکاران (۲۰۱۱) هم‌واژگانی را یکی از روش‌های کتاب‌سنجی و علم‌سنجی معرفی می‌کنند. تاکنون، در پژوهش‌های متعددی در زبان فارسی، روش‌های علم‌سنجی از جمله زوج‌های کتاب‌شناختی، تحلیل‌های هم‌استنادی، و تحلیل‌های روابط هم‌نویسندگی تبیین شده است؛ اما تحلیل هم‌واژگانی کمتر از سایر روش‌های علم‌سنجی به کار گرفته شده است. به‌همین دلیل، در مقاله حاضر تاریخچه و پیشینه، کاربردها، الگوریتم‌ها، روش‌شناسی، مزایا، و ضعف‌های تحلیل هم‌واژگانی بیان شده است.

تعاریف

هم‌واژگانی را از نظر مفهومی معادل co-word دانسته که گاه واژه co-occurrence را به‌جای آن به‌کار برده‌اند. در زبان فارسی واژه نخست را هم‌واژگانی و واژه دوم را هم‌رخدادی معنا کرده‌اند. تحلیل هم‌واژگانی که براساس هم‌رخدادی واژگان عمل می‌کند، به‌عنوان یک روش تحلیل محتوا، یکی از روش‌های علم‌سنجی است (کینگ^۱، ۱۹۸۷). گاه تحلیل هم‌واژگانی به‌عنوان جانشینی برای تحلیل هم‌استنادی در نظر گرفته می‌شود (لیدسدورف، ۱۹۹۱). تحلیل هم‌واژگانی نمونه‌ای از روش مدل‌سازی گرافیکی است که در آن از اندیشه‌های مربوط به تحلیل رابطه استفاده می‌شود (نف و کرلی^۲، ۲۰۰۹). این روش نیز ابزاری قدرتمند در کشف دانش و ترسیم نقشه کتاب‌شناختی است.

علاوه بر اینها، تعریف‌های دیگری توسط برخی پژوهشگران از جمله کورتیال^۳، ۱۹۹۴؛ ادواردز^۴، ۱۹۹۵؛ دینگ، چاودری، و فو^۵، ۲۰۰۱؛ روکایا^۶، ۲۰۰۸؛ کوستوف^۷، ۱۹۹۸؛ و هی^۸، ۱۹۹۹ ارائه شده است که در زیر به آن اشاره شده است:

۱. نوعی تحلیل محتوا برای کشف الگوها و تعیین گرایش‌های موضوعی در حوزه‌های

مشخص پژوهشی؛

1. King
2. Neff & Corley
3. Courtial
4. Edwards
5. Ding, Chowdhury, & Foo
6. Rokaya
7. Kostoff
8. He

۲. اندازه‌گیری درجه ارتباط بین مفاهیم و اصطلاحات؛
 ۳. روشی مهم برای دیداری‌سازی روابط میان مفاهیم، اندیشه‌ها، و مسائل علمی؛
 ۴. تحلیلی کمی برای کشف شبکه مفاهیم حوزه‌های علمی؛
 ۵. روشی برای شناسایی ساختار موضوعی یک حوزه پژوهشی؛
 ۶. کشف روابط مفهومی میان مدارک؛
 ۷. شیوه‌ای نو از مطالعات کمی در حوزه علم‌سنجی که از طریق آن ارتباط میان توصیفگرهای منابع منتشر شده فراهم می‌شود؛ و
 ۸. به‌کارگیری این روش سبب شناسایی و دسته‌بندی مقولات مفهومی می‌شود و زمینه را برای بازبینی اطلاعات در پایگاه‌های دانش فراهم می‌آورد.
- با توجه به تعاریف فوق، تحلیل هم‌واژگانی شیوه‌ای از تحلیل محتواست که از طریق هم‌رخدادی واژه‌ها یا مفاهیم موجود در متون و منابع حاصل می‌شود و از طریق آن می‌توان مفاهیم اصلی یک زمینه یا حوزه علمی را شناخت و به‌واسطه این شناخت، الگوها و رویدادهای مفهومی، ساختار علمی، شبکه مفهومی، روابط سلسله‌مراتبی مفاهیم، و مقولات مفهومی آن حوزه را کشف، ترسیم، و مدیریت کرد. تحلیل هم‌واژگانی، ابزاری برای کشف الگوهای پنهان و رویدادهای نوظهور مفهومی است.

کارکردها

مطالعات مختلف در حوزه علم اطلاعات، به‌خصوص در زمینه تحلیل هم‌واژگانی، نشان می‌دهد این روش می‌تواند در موارد زیر کاربرد داشته باشد:

۱. ترسیم حرکت و پویایی علم (کالون، کورتیال، و ترنر، ۱۹۸۶)؛
۲. ترسیم ساختار حوزه‌ها و زمینه‌های علمی (وایتاگر، ۱۹۸۹)؛
۳. ترسیم روابط میان پژوهش‌های بنیانی و پژوهش‌های فناورانه (کالون و همکاران، ۱۹۹۱)؛
۴. دیداری‌سازی شبکه مفهومی علم و فناوری (سالمی و کوشا، ۱۳۹۲)؛
۵. تحلیل سیر تحول مفهومی در طول دوره‌های زمانی؛
۶. تعیین و تحلیل نواحی مورد پژوهش؛
۷. ارزیابی درون‌داد و برون‌داد روابط در یک شبکه پژوهشی (عصاره و همکاران، ۱۳۹۴، نقل از ترنر و روژان، ۱۹۹۱)؛
۸. دسته‌بندی مدارک براساس موضوعات؛
۹. خوشه‌بندی مفاهیم حوزه‌های علمی؛

1. Collan, Courtial, & Turner
2. Whittaker
3. Turner & Rojouan

۱۰. برنامه‌ریزی و سیاست‌گذاری در حوزه‌های علمی؛
۱۱. درک و کشف روابط پنهان در حوزه‌های علوم مختلف؛
۱۲. کشف الگوهای برجسته و رویدادهای در حال ظهور؛ و
۱۳. تعیین روابط سلسله‌مراتبی مفاهیم یک حوزه یا زمینه خاص علمی.

پیشینه تحلیل هم‌واژگانی

نخستین کوشش‌ها در زمینه واژگان درون‌متنی توسط زیف^۱ در سال ۱۹۴۹ صورت گرفت. زیف با مطالعه واژه‌هایی که در متون انگلیسی به کار رفته‌اند به مصادیقی برای اصل کمترین کوشش دست یافت. تحلیل بسامد واژگان در عناوین مقالات، ابتدا توسط اسمال و گریفث در سال ۱۹۷۴ به کار رفت تا موضوع یا به‌طور خاص، مفهوم کلی مدرک استنادشده را مشخص کنند.

تحلیل هم‌واژگانی نخستین بار در دهه ۱۹۸۰ در فرانسه در مرکز جامعه‌شناسی خلاقیت کول^۲ به کار گرفته شد. این تحلیل به‌عنوان شاخصی در علم‌سنجی در این دهه معرفی شد و بعد از آن توسعه پیدا کرد و اغلب به‌عنوان جایگزینی برای رویکردهای استنادی و هم‌استنادی برای ترسیم علم استفاده شد.

روش تحلیل هم‌واژگانی در سال‌های اخیر در پژوهش‌های پژوهشگران کشورهای مختلف به‌نحو قابل ملاحظه‌ای مورد توجه قرار گرفته است. بسیاری از پژوهشگران با این روش حوزه‌های خاصی را به قصد تحلیل و ترسیم ساختار آن مطالعه کرده‌اند. نمونه‌هایی از به‌کارگیری این روش توسط پژوهشگران زیر مشاهده می‌شود: هی، ۱۹۹۹؛ دینگ، چاودری، و فو، ۲۰۰۱؛ بین^۳ و همکاران، ۲۰۰۹؛ لیدسدورف، ۲۰۱۰؛ میلوجویک، سایموتو، و یان^۴، ۲۰۱۱؛ لیو، هو، و وانگ^۵، ۲۰۱۲؛ یانگ، هو، و سیو^۶، ۲۰۱۲؛ هو، دنگ، و لیو^۷، ۲۰۱۳؛ زونگ^۸ و همکاران، ۲۰۱۳؛ راوی کومار، اگرهاری، و سینگ^۹، ۲۰۱۴.

در ایران پژوهش مستقلی در مباحث نظری روش تحلیل هم‌واژگانی انجام نگرفته است، اما با استفاده از کاربردهای این روش، حوزه‌های علمی متعددی مطالعه و تحلیل شده است. در این میان، می‌توان به ابوبی اردکانی و همکاران (۱۳۸۸)؛ توکلی‌زاده اوروی و نجابتیان (۱۳۸۹)؛ ناصری‌جزه، طباطبائی‌ان، و فاتح‌راد (۱۳۹۱)؛ الهی و همکاران (۱۳۹۱)؛ احمدی، سلیمی، و زنگی‌ش (۱۳۹۲)؛ سالمی و کوشا (۱۳۹۲)؛ صدیقی (۱۳۹۳)؛ کاتبی (۱۳۹۳)؛ احمدی و کوبی (۱۳۹۴)؛ و ذوالفقاری و همکاران (۱۳۹۴) اشاره کرد. وجه مشترک این پژوهش‌ها روش تحلیل هم‌واژگانی برای ترسیم ساختار حوزه‌های مورد مطالعه است. همچنین، در بخش روش‌شناسی پژوهش‌های ذکرشده، نکات مفید و فنی درباره این روش بیان شده و وجه

1. Zipf
2. Center De Sociologie de La Innovation of the Ecole
3. Yin
4. Milojevic, Sugimoto, & Yan
5. Liu, Hu, & Wang
6. Yang, Wu, & Cui
7. Hu, Deng, & Liu
8. Zong
9. Ravikumar, Agrahari & Singh

مشترک آن استفاده از بسامد واژگان در تحلیل این روش است. در همه این پژوهش‌ها اشاره مختصری به کاربردهای این روش شده و از آن فراتر نرفته است.

پیش‌فرض‌های تحلیل هم‌واژگانی

تحلیل هم‌واژگانی که در سال ۱۹۸۳ توسط کالون مطرح شد، بر این فرض استوار بود که حضور واژه‌ها یا مفاهیم در کنار یکدیگر نشان‌دهنده محتوای آن مدرک است. بنابراین، با اندازه‌گیری میزان این هم‌رخدادی می‌توان شبکه مفاهیم یک زمینه علمی را ترسیم کرد. این شبکه مفهومی با شمارش تعداد دفعات حضور هر اصطلاح موضوعی و هم‌نشینی آن با موضوعات دیگر ترسیم می‌شود. به عبارتی، اگر دو اصطلاح با هم در یک مدرک به کار روند و هر چه بیشتر با هم تکرار شوند، یعنی این دو واژه ارتباط معنایی بیشتری با هم دارند. از هم‌رخدادی دو اصطلاح یا دو واژه برای کشف پیوند و رابطه میان دو موضوع در یک حوزه پژوهشی نیز استفاده می‌شود و از این طریق می‌توان توسعه و پیشرفت آن حوزه از علم را ردیابی کرد.

این رویکرد خود بر سه فرضیه استوار است: (۱) واژگان مورد استفاده در متون علمی به‌دقت توسط نویسندگان انتخاب می‌شوند، (۲) استفاده از واژگان مختلف در یک متن، نشان‌دهنده وجود برخی روابط غیرجزئی میان آنهاست، و (۳) تکرار هم‌رخدادی‌های واژه‌ها توسط نویسندگان مختلف در متون مختلف به این معناست که روابط میان این واژه‌ها در حوزه علمی مورد مطالعه حائز اهمیت است (عصاره و همکاران، ۱۳۹۴، نقل از میلوچویک، ۲۰۰۹).

ویتاکر (۱۹۸۹) و هی (۱۹۹۹) نیز بر پیش‌فرض‌های زیر تأکید ورزیده‌اند:

- پدیدآورندگان در انتخاب اصطلاحات موضوعی نهایت دقت را دارند؛
- به‌کارگیری واژه‌ها در مدارک نشان‌دهنده ارتباط مستقیم آن با محتوای مدارک است؛
- واژگان داخل متن روابط معنایی موضوع یا حوزه مورد نظر را نشان می‌دهند؛ و
- کلیدواژه‌های توصیفی نمایه‌سازان منبع خوبی برای تحلیل هم‌واژگانی است.

تفاوت تحلیل هم‌واژگانی با تحلیل هم‌استنادی^۱

تفاوت مهم میان تحلیل هم‌واژگانی و هم‌استنادی این است که تحلیل هم‌استنادی در دوره‌های پژوهشی مشخص به منابع استناددهنده و مآخذ استنادشده نیاز دارد، اما تحلیل هم‌واژگانی فقط نیازمند مجموعه‌ای از مقالات مجلات در حوزه موضوعی خاصی است. داده‌های حاصل از این روش توسط ماتریس بسامد سنجیده می‌شود و می‌توان نتایج حاصل را همچون تحلیل

1. Co-citation analysis

هم‌استنادی در خوشه‌های سلسله‌مراتبی یا مقیاس چندبعدی به‌نمایش گذاشت. تحلیل هم‌واژگانی را ابزار مناسب‌تری برای مطالعه روابط درونی میان پژوهش‌های دانشگاهی و فناوریانه دانسته‌اند، زیرا شاخص‌های تحلیل هم‌استنادی فقط وجود ارتباط را نشان می‌دهد و اطلاعاتی از موضوع مورد پرسش را ارائه نمی‌دهد. برای درک این مسئله که آیا پژوهش‌های علمی عامل اصلی اختراع و ابداع است یا فناوری، باید به خود متون مراجعه و محتوای مدارک و اختراعات را بررسی کرد. این کار با تحلیل واژگان داخل متون فراهم می‌شود.

منابع اطلاعاتی در تحلیل هم‌واژگانی

دامنه واژه‌های مورد استفاده در تحلیل هم‌واژگانی از کلیدواژه‌های یک اصطلاحنامه تا واژه‌های داخل یک متن کامل گسترده است. نخستین مطالعات تحلیل هم‌واژگانی، براساس کلیدواژه‌های اصطلاحنامه‌ای انجام شده است (هی، ۱۹۹۹، نقل از باین^۱ و دیگران، ۱۹۸۶). سپس متون براساس عنوان، خلاصه، یا تعداد مشخصی از کلیدواژه‌ها منحصراً به توصیف‌گرهای یک اصطلاحنامه مطالعه شدند (کالون و همکاران، ۱۹۹۱). روتو و مورگان^۲ (۱۹۹۷) پیشنهاد دادند که تحلیل هم‌واژگانی را می‌توان در سطح چکیده و با واژه‌های پیشنهادی متخصصان اجرا کرد. کوستوف، ابرهارت، و تویتمن^۳ (۱۹۹۷) معتقد بودند یکی از امتیازات تحلیل تمام‌متن، توانایی بازیابی عبارات مهم و در عین حال کم‌بسامد است که در سایر تحلیل‌ها نادیده گرفته می‌شود.

نویون و ون‌ران^۴ (۱۹۹۸) فرایند انتخاب واژگان را نشانه DNA یک مدرک قلمداد کرده‌اند. در تمام انتشارات، کلیدواژه‌ها آگاهانه انتخاب می‌شوند و عبارتی را تشکیل می‌دهند تا ویژگی‌های یک رشته را توصیف کنند. به سخن دیگر، مجموعه هزاران مدرک یک حوزه پژوهشی شبکه بزرگی را تشکیل می‌دهند که در آن، کلیدواژه‌ها مدارک را به هم پیوند می‌دهند. هرچه واژه‌های مشترک دو مقاله بیشتر باشد این مقالات ارتباط نزدیک‌تری دارند و در نتیجه دارای مشابهت موضوعی هستند. در یک استعاره بیولوژیکی می‌توان گفت هرچه DNA دو مقاله به هم شبیه‌تر باشند ارتباط آنها نزدیک‌تر است. اگر این شباهت از حد مشخصی بیشتر شود آنها به یک حوزه خاص پژوهشی متعلق هستند. برای بازنمایی این شبکه‌های انتشاراتی از فنون ریاضی استفاده می‌کنند. شباهت واژگانی خوشه‌بندی و نقشه‌دووجهی از خوشه‌ها ابزارهایی هستند که این فرایند ساختار پنهان را به تصویر می‌کشند. این ساختارها را می‌توان ساختار شناختی حوزه علمی دانست.

کیفیت نتایج حاصل از تحلیل هم‌واژگانی به عوامل زیادی مانند کیفیت کلیدواژه‌ها و

1. Bauin
2. Roto & Morgan
3. Koštoff, Eberhart & Toothman
4. Noyons & Van Raan

واژه‌های نمایه‌ای، دامنه پایگاه اطلاعاتی، قابلیت روش‌های آماری و ارائه یافته‌ها بستگی دارد. مهم‌ترین مسئله در این روش، انتخاب کلیدواژه است که از آن به "تأثیر نمایه‌ساز" یاد شده است و بیشتر پژوهشگران بر آن تأکید کرده‌اند. نمایه‌سازی مداخله تحلیل‌گر با متن است و اعتبار نقشه به ماهیت نمایه‌سازی بستگی دارد. ویتاگر (۱۹۸۹) معتقد است که نتایج تحلیل هم‌واژگانی به نحوه انتخاب کلیدواژه‌ها توسط نمایه‌ساز بستگی دارد تا از طریق آنها حوزه‌های علمی را مفهوم‌سازی کند.

روش‌شناسی تحلیل هم‌واژگانی

تحلیل هم‌واژگانی دارای مراحل اجرایی خاصی است که تلویحاً می‌توان به صورت گام‌های زیر تعریف کرد:

گام اول: انتخاب حوزه یا زمینه مورد مطالعه. با توجه به کارکردهای این روش، انتخاب حوزه یا زمینه مستعد مطالعه مهم است. از آنجا که منابع مورد نیاز برای این تحلیل، واژگان یا مفاهیم هستند، اگر انتخاب زمینه و حوزه‌ای که دارای مفاهیم غنی یا از لحاظ جامعیت و مانعیت مفهومی قابل طرح باشد نتایج دقیق‌تری حاصل می‌شود. بنابراین، انتخاب حوزه یا زمینه مورد مطالعه با این روش باید با محدودیت‌ها و کارکردهای این روش هم‌خوان باشد.

گام دوم: استخراج مفاهیم یا واژگان از مدارک. برای استخراج مفاهیم از مدارک دو شیوه وجود دارد: شیوه اول، به روش تحلیل محتوا و به صورت دستی و شیوه دوم با روش‌های متن‌کاوی و به صورت خودکار انجام می‌شود. در شیوه اول، واژگان و مفاهیم از عنوان، چکیده، و کلیدواژگان توصیفگر یک‌جا استخراج می‌شود. واژگان عنوان و چکیده مقالات، نشانه‌های ساختار یک مدرک علمی هستند که از دو نظر با هم مشابه‌اند. اول اینکه آنها عبارات یا واژه‌های مرتبط با محتوای مقاله علمی هستند و ابزار خلاصه‌سازی، چکیده‌سازی، و یا طبقه‌بندی موضوعات مقاله‌اند. دوم، این واژه‌ها نشانگرهای شناسایی و ممتاز کردن هر مدرک هستند. لیدسدورف (۱۹۸۷، ۱۹۸۹) به تأثیر نمایه‌ساز در تحلیل‌های هم‌واژگانی انتقاد کرد و پیشنهاد داد که برای رفع این اشکال باید از واژه‌های عنوان برای این تحلیل استفاده کرد. این روش، دسترسی مستقیم به نظرات مؤلفان را امکان‌پذیر می‌کند و توصیفگرها را قابل اعتمادتر از واژه‌های نمایه‌ساز می‌داند. ویتاگر (۱۹۸۹) برخلاف لیدسدورف، معتقد بود که مؤلفان واژه‌ها را برای تأثیر بر مخاطبان انتخاب می‌کنند و از سوی دیگر همه عناوین استاندارد نیستند.

در تحلیل هم‌واژگانی کلیدواژه‌ها در عنوان، چکیده، یا متن مقالات بررسی می‌شود. این روش با تحلیل مجموعه‌ای از مدارک به ارزیابی میزان ارتباط آنها می‌پردازد. در این تحلیل برای واحد اندازه‌گیری تعداد واژه‌ها، هم‌رخدادی کلیدواژه‌ها را در عنوان، چکیده، یا متن مقالات بررسی می‌کنند. متن‌ها به‌عنوان واحد تحلیل در نظر گرفته می‌شوند. این متن‌ها از لحاظ اندازه از متن طولانی گرفته تا تعداد جملات کوتاه، همچون عنوان متن، متغیرند. متن‌ها شامل واژه‌هایی هستند که می‌توان آنها را به‌صورت جمله، پاراگراف، و بخش سازماندهی کرد. ساختار معنایی در روابط بین واژگان در رده‌های مختلف ممکن است متفاوت باشد (لیدسدورف و ولبرز، ۲۰۱۱).

پژوهشگران در خصوص این شیوه رفتارهای متفاوتی داشته‌اند. برای نمونه کامبروسو، لیموز، کورتیال، و لایویل^۲ (۱۹۹۳) از کلیدواژگان عناوین و کلیدواژگان نمایه‌سازی استفاده کردند و کالتر، مونارچ، و کانداس^۳ (۱۹۹۸) توصیفگرهای نمایه‌سازان حرفه‌ای را انتخاب کردند. دیلوز و لماریا^۴ (۱۹۹۷) از کلیدواژه‌های متخصصان موضوعی و نیویون و ونران (۱۹۹۸) از کدهای سازماندهی شده پایگاه‌های اطلاعاتی استفاده کردند. بنابراین، عنوان و چکیده و تا حدودی توصیفگرهای مدرک که توسط نویسنده به‌عنوان کلیدواژه مدرک توصیف می‌شود بیشترین نقاط مورد استخراج مفاهیم یا کلیدواژه‌ها در شیوه نخست است.

شیوه دوم، استخراج مفاهیم به‌طور مستقیم از متن مدارک است. این شیوه به‌دلیل گسترده بودن متن‌ها، دشوار و زمان‌بر است و هزینه آن هم بالاست. در این شیوه، استخراج معمولاً با روش متن‌کاوی به‌صورت خودکار انجام می‌شود.

گام سوم: آماده‌سازی و انتخاب نهایی مفاهیم. مرحله آماده‌سازی و انتخاب نهایی

واژگان یا مفاهیم از اهمیت ویژه‌ای برخوردار است و به‌دلیل مواردی که مطرح می‌شود اگر با دقت انجام نگیرد در مراحل بعدی و در نتیجه‌گیری، پژوهش دچار نقص خواهد شد. در این مرحله، معمولاً چهار عمل صورت می‌گیرد شامل یکدست‌سازی واژه‌ها، برچسب‌گذاری، تهیه سیاهه کلمات بازدارنده، و انتخاب روش و عملی که به‌وسیله آن از میان واژه‌های استخراج‌شده اولیه، مفاهیم یا کلیدواژه‌های مهم انتخاب شود. اجرای کارهای فوق آسان نیست. زیرا هرکدام تابع قواعد خاصی است و باید با دقت انجام شود، به‌طور مثال، استفاده از پالایه‌های زبان‌شناسی در ترکیب‌بندی واژگان، مسائل دستوری، حروف اضافه، و به‌کارگیری شکل جمع یا مفرد واژگان.

بنابراین، پژوهشگران باید ابتدا تعیین کنند که چه چیزی را به‌عنوان واحدهای مرتبط با تحلیل در نظر می‌گیرند. دوم اینکه چه کلماتی را باید در تحلیل وارد ساخت. در این مقاله

1. Welbers
2. Cambrosio, Limoges, Courtial, & Laville
3. Coulter, Monarch & Konda
4. De Looze & Lemarié

به دو روشی پرداخته می‌شود که سبب انتخاب و استخراج مفاهیم بااهمیت می‌گردد. یکی از آنها استفاده از روش C-value و دیگری عمل وزن‌دهی به مفاهیم است. روش C-value ترکیبی از روش‌های زبانی و آماری را برای استخراج واژه‌های ترکیبی استفاده می‌کند (فرانتزی، آنانیودیو، و می‌ما، ۲۰۰۰). روش زبانی C-value شامل برچسب‌گذاری بخشی از کلام، پالایه زبانی، و سیاهه کلمات توقف است. برچسب‌گذاری بخشی از کلام، اختصاص دادن برچسب‌های گرامری (نظیر اسم، صفت، فعل، حرف اضافه، ضمیر، و قید) به هر واژه در متن است. روش آماری C-value برای استخراج اصطلاحات چندکلمه‌ای به کار می‌رود و هدف آن بهبود استخراج واژه‌های طولانی‌تر است که ممکن است به‌تنهایی در متن رخ ندهند. پالایه زبانی (اسم + اسم و صفت + اسم) بهترین روش برای انتخاب کلیدواژگان باکیفیت است. به‌منظور محاسبه مقدار C-value رشته a دو حالت زیر در نظر گرفته می‌شود:

الف. اگر a رشته‌ای با بیشترین طول باشد یا تودرتو نباشد، مقدار C-value با استفاده از فراوانی کلی آن در متن و طول آن براساس رابطه زیر به‌دست می‌آید. |a| طول رشته a و f(a) فراوانی رخداد آن در متن است:

$$C - value(a) = \log_2 |a| \cdot (f(a))$$

ب. اگر a رشته‌ای تودرتو باشد باید بررسی شود که آیا بخشی از واژه‌ها با طول بلندتر است. اگر چنین باشد برای محاسبه مقدار C-value باید فراوانی آن به‌عنوان یک رشته تودرتو و تعداد واژه‌های طولانی‌تر محاسبه شود. در این حالت، مقدار C-value براساس رابطه زیر محاسبه می‌شود:

$$C - value(a) = \log_2 |a| \cdot (f(a) - \frac{1}{P(T_a)} \sum_{b \in T_a} f(b))$$

|a| طول رشته a، f(a) فراوانی رخداد رشته a در متن، T_a مجموعه رشته‌های نامزد استخراج‌شده حاوی a، $P(T_a)$ تعداد عناصر T_a و $\sum_{b \in T_a} f(b)$ مجموع فراوانی‌هایی است که در رشته‌های طولانی‌تر رخ می‌دهد (فرانتزی، آنانیودیو، و می‌ما، ۲۰۰۰).
به‌منظور درک بهتر نحوه محاسبه در زیر به نمونه‌هایی اشاره می‌شود:
اگر دو مفهوم مانند "ضریب تأثیر" با تعداد رخداد ۲۰ و "ضریب تأثیر مجلات" با رخداد ۸ داشته باشیم، ارزش C-value آنها براساس محاسبه فرمول فوق به‌ترتیب ۱۱/۰۶ و ۱۶/۵ است.

$$\log_2 |3| \cdot 8 = 11.06 \quad , \quad \log_2 |2| \cdot (20 - \frac{1}{2} (8)) = 16.5$$

1. Frantzi, Ananiadou, & Mima
2. Frantzi, Ananiadou, & Mima

با این حال، سالتون و مک‌گیل^۱ (۱۹۸۳) معتقدند که واژه‌های با بیشترین و کمترین فراوانی ممکن است کم‌اهمیت‌تر از واژه‌های با فراوانی متوسط باشند. به این منظور، نویسندگان شاخصی را با عنوان روش معمول TF-IDF یا "فراوانی واژه-معکوس فراوانی سند" پیشنهاد کردند.

TF بیانگر نسبت مدارک دربرگیرنده آن مفهوم در بین تمامی مدارک است. برای محاسبه آنها ابتدا فراوانی واژه i در مدرک j (F_{ij}) محاسبه می‌شود و با هنجار کردن آن در تمامی مجموعه، مقدار TF به دست می‌آید.

$$TF_{ij} = F_{ij} / \max (F_{ij})$$

$$IDF_i = \log (N/n_i)$$

IDF بر این اساس است که واژه‌هایی که در مدارک فراوانی ظاهر می‌شوند، بیشتر بیانگر موضوع خاص هستند. به همین دلیل، برای محاسبه آن ابتدا تعداد اسنادی که دربرگیرنده واژه i هستند (n_i) و تعداد کل اسناد در مجموعه (N) را مشخص می‌کنند، سپس IDF از رابطه زیر محاسبه می‌شود:

$$TF - IDF = TF_{ij} * IDF_i$$

این رابطه نشان‌دهنده حاصل ضرب TF در IDF و بیانگر اهمیت یک واژه یا مفهوم در مدرک است و می‌توان براساس آن مفاهیم موجود در مدارک را برحسب میزان اهمیت آنها رتبه‌بندی کرد. برای درک بیشتر، چنانچه مفهوم "اخلاق علمی" با بسامد ۴۲ در حوزه‌ای که یکی از مفاهیم آن با بالاترین بسامد آمده باشد (۶۰)، وزن مفهوم "اخلاق علمی" نسبت به کل حوزه بدین شرح محاسبه می‌شود: در این حوزه فرضی ۲۰ مدرک درباره اخلاق علمی آمده است. مطابق فرمول اول $0.7 = 42 \div 60$. مطابق فرمول دوم لگاریتم کل مدارک حوزه فرضی (۷۴۰) تقسیم بر تعداد مدارک اخلاق علمی (۲۰) مساوی $1/57$. مطابق فرمول سوم $1/0.9 = 1/57 \times 0.7$. به این صورت، وزن مفهوم اخلاق علمی تقریباً برابر ۱ است. به این ترتیب، با انتخاب یک آستانه مورد نظر می‌توان با انجام محاسبات فوق در یکی از دو روش مذکور، مفاهیم یا واژگان بااهمیت را در تحلیل هم‌واژگان انتخاب کرد و نتیجه بهتری به دست آورد. علاوه بر آن، معمولاً از روش وزن‌دهی برای تعیین خوشه‌های مفهومی در تحلیل هم‌واژگانی استفاده می‌شود.

به‌طور خلاصه، پژوهشگر می‌تواند از چهار شاخص برای انتخاب سیاهه واژگان دخیل

1. Salton & McGill

در تحلیل استفاده کند: ۱) فراوانی واژه، ۲) مقدار $fDI-FT$ (۳) مشارکت ستون‌ها با ماتریس، و ۴) مجموع تفاضل مقدار مشاهده‌شده/ مقدار پیش‌بینی شده برای هر کلمه. در مطالعه‌ای که توسط لیدسدورف و ولبرز (۱۱۰۲) صورت گرفت استفاده از شاخص اخیر آسان‌تر بوده است، هر چند که هر چهار روش قابل سنجش هستند.

گام چهارم: ساخت ماتریس هم‌رخدادی واژگان. تحلیل ویژگی‌ها و شاخص‌های

ماتریس هم‌رخدادی مهم‌ترین مرحله از تحلیل هم‌رخدادی واژگان است. در بیشتر موارد، دو هدف مورد نظر است: یکی شناسایی سلسله‌مراتبی مفاهیم در میان حوزه‌های یک مسئله پژوهشی، دوم شناسایی حوزه‌های کوچک اما مستعد رشد. کالون، کورتیال، و ترنر (۱۹۸۶) برای پاسخ به این دو درخواست دو شاخص معرفی می‌کنند: شاخص شمول^۱ و شاخص مجاورت^۲. ماتریس پایه بیانگر تکرار واژگان در متن است. با توجه به دو هدف فوق در این تحلیل دو نوع ماتریس مورد نیاز است. ماتریس "واژه-واژه" و ماتریس "واژه-سند". معمولاً در تحلیل هم‌واژگانی، از این دو ماتریس برای تعیین روابط سلسله‌مراتبی مفاهیم استفاده می‌شود.

ارزش هر سلول در ماتریس براساس تعداد دفعاتی تعیین می‌شود که دو واژه یا مفهوم در کنار یکدیگر در یک مدرک ظاهر می‌شوند. در این حالت، بالاترین میزان رخداد دو واژه در ماتریس به معنای نزدیک‌ترین ارتباط بین دو واژه است (روکویا، ۲۰۰۸). ماتریس "واژه-واژه" از نوع متقارن است؛ یعنی سطرها و ستون‌ها برابر هستند؛ اما ماتریس "واژه-سند" می‌تواند نامتقارن باشد و ستون‌ها، سندها، یا مدارک و سطرها مفاهیم این ماتریس را تشکیل می‌دهد. در این ماتریس، متون برای بیان حالاتی (مثل ردیف‌ها) در نظر گرفته می‌شوند که در آن واژگان به‌عنوان متغیر (ستون‌ها) به آنها نسبت داده می‌شوند (لیدسدورف و ولبرز، ۲۰۱۱). هریک از این دو ماتریس با هدف‌های خاصی ساخته می‌شوند. معمولاً برای ایجاد شبکه مفهومی یا واژگانی از ماتریس متقارن و برای استخراج روابط سلسله‌مراتبی مفاهیم از ماتریس نامتقارن استفاده می‌شود.

گام پنجم: تحلیل و ترسیم نقشه‌های مفهومی با داده‌های هم‌رخدادی. ترسیم نقشه‌های

مفهومی یکی از خروجی‌های داده‌های هم‌رخدادی واژگان است که معمولاً بعد از مرحله ماتریس‌سازی انجام می‌شود. ایجاد نقشه‌های مفهومی نوعی ترسیم ساختار حوزه تلقی می‌شود. هر حوزه علمی، زمینه‌ای از ساختار دانش آن حوزه به حساب می‌آید. ترسیم نقشه را گامی برای صرفه‌جویی در زمان و هزینه در پیاده‌سازی انواع ابتکارات مدیریت علم و

1. Inclusion index
2. Proximity index
3. Rokaya

پژوهش و دستیابی به مزایای بلندمدت می‌دانند. مطالعه نقشه‌های علمی یکی از مهم‌ترین وجوه مطالعات علم‌سنجی است که در بسیاری از حوزه‌های دیگر مورد توجه است. نقشه‌های علم با ترسیم گرافیکی هر رشته علمی راه را برای شناسایی هرچه بهتر و دقیق‌تر آن شاخه از دانش بشری و تبدیل مفهوم انتزاعی رشته علمی به مفهومی عینی تر هموار کرده‌اند (آنگون^۱ و همکاران، ۲۰۰۵). این نقشه با فنون و روش‌های متعددی ترسیم می‌شود. یکی از اهداف و کاربردهای تحلیل هم‌واژگانی، ترسیم ساختار علم یا همان ترسیم نقشه‌های علمی است. کیفیت نقشه‌ها به تعداد مدارک، تعداد مفاهیم یا کلیدواژه‌ها، و نوع ماتریس و نرم‌افزار بستگی دارد. برای تحلیل نقشه‌های ترسیم‌شده مهارت فنی و آگاهی از حوزه مورد نظر ضروری است.

تحلیل نگاشت نقشه علمی را می‌توان با استفاده از نرم‌افزارهای عمومی تحلیل شبکه اجتماعی انجام داد؛ اما نرم‌افزارهای دیگری وجود دارد که به صورت تخصصی برای تحلیل نگاشت علمی ایجاد شده‌اند. با پیشرفت نرم‌افزارهای رایانه در حوزه علم‌سنجی، ترسیم و تحلیل نقشه‌های علمی آسان‌تر شده است و با دقت زیادی انجام می‌گیرد. در سال‌های اخیر، نرم‌افزارهای متعددی در ترسیم و تحلیل داده‌های شبکه‌ای به خصوص در تحلیل هم‌واژگانی به کار گرفته شده است، مانند: IN-SPIRE، ColPalRed، CiteSpaceII، Pajek، Ucinet، Bibexcel، Google، Vantage Point، Sci Tool، Network Workbench، Fulltext.exe، NetDraw، VOSViewer و Google Earth، Maps، nodexl. بعضی از این نرم‌افزارها صرفاً برای تجزیه و تحلیل داده و ایجاد ماتریس هم‌رخدادی به کار گرفته می‌شوند و کاربردی در ایجاد نگاشت نقشه ندارند.

الگوریتم‌های تحلیل هم‌واژگانی

الگوریتم تحلیل هم‌واژگانی برای تولید شبکه‌های مطلوب با استفاده از داده‌ها از دو روش استفاده می‌کند: روش نخست، شبکه‌ای را تشکیل می‌دهد که قوی‌ترین ارتباطات موجود بین واژه‌ها را به تصویر می‌کشد و پیوندهایی که در این روش ایجاد می‌شوند "پیوند درونی" نام دارند. روش دوم، پیوندهای ضعیف‌تر بین واژگان موجود در شبکه‌ها را نیز نشان می‌دهد. پیوندهایی که در روش دوم ایجاد می‌شوند "پیوند بیرونی" نام دارند (کافمن و روسو، ۱۹۹۰). دو واژه کلیدی را هم‌ایند یا هم‌واژگانی گویند که در متن‌های ثبت‌شده و در پایگاه داده با هم نمود یافته باشند. فرض کنید C_{ij} تعداد رخداد‌های واژه کلیدی i در کل متون یک مجموعه داده باشد. اگر C_{ij} تعداد رخداد‌های واژگان کلیدی i و j باشد، میزان پیوستگی و ارتباط بین دو توصیفگر i و j از طریق رابطه زیر به دست می‌آید: S : عدد صحیح X عدد صحیح تابع حقیقی $S(C_{ij}, c_i, c_j) = :$ اگر $(c_i \leq 0) \vee (c_j \leq 0)$ آنگاه $S = :$ تعریف نشده؛ در

1. Anegón

غیر این صورت $S \leftarrow$ در بازه‌ای بین $0 \leq S \leq 1$ قرار خواهد داشت:

$$\frac{c_{ij}^2}{c_i c_j} = : S$$

اگر تابع ارتباط بین واژگان کلیدی به ۱ نزدیک‌تر باشد پیوند بین گره‌ها در شبکه را نشان می‌دهد و این شبکه قوی‌ترین ارتباطات موجود در مجموعه داده‌ها را به تصویر می‌کشد. عدد نشان‌دهنده میزان ارتباط بین دو واژه که به‌ندرت در یک مجموعه داده با هم می‌آیند اما همیشه با هم هستند بزرگ‌تر از عدد مربوط به واژگان کلیدی است که بارها در مجموعه داده تکرار می‌شوند و تقریباً همیشه با هم می‌آیند. از این رو، احتمال دارد که ارتباطات ضعیف یا بی‌ربط در شبکه غالب باشد. یک راه برای حل این مسئله این است که فقط آن دسته از واژگان کلیدی را به‌عنوان عامل پیوندی در ساختن شبکه در روش اول در نظر بگیریم که دفعات هم‌رخدادی آنها فراتر از مقدار حداقل باشد (کولتر، مونارک، و کوندا، ۱۹۹۸).

در روش نخست، پیوندی که بزرگ‌ترین عدد هم‌آیندی را داشته باشد اول انتخاب می‌شود و گره آن به‌عنوان گره آغازین در شبکه در نظر گرفته می‌شود. پیوندهای دیگر و گره‌های متناظر آنها نیز از طریق بررسی میزان ارتباط پیوندها به گراف اضافه می‌شوند. یعنی قوی‌ترین پیوند، گره‌ای را که در هیچ گرافی به‌کار نرفته به گرافی پیوند می‌زند که در حال شکل‌گیری است و قبل از همه به شبکه افزوده شده است.

شبکه بعدی به‌روش مشابهی تولید می‌شود شکل‌گیری آن با پیوند با بزرگ‌ترین عدد هم‌آیندی آغاز می‌شود که در گراف موجود نیست. در روش دوم الگوریتم، با انتخاب پیوندهایی که از بزرگ‌ترین میزان هم‌رخدادی برخوردارند و از آستانه هم‌واژگانی فراتر هستند و در بعضی از شبکه‌های روش یک وجود دارند به هریک از گراف‌های موجود گره‌هایی اضافه می‌شود. شبکه‌های تولیدشده توسط ابزار تحلیلی هم‌آیند، توزیع مفاهیم مربوط به مرکزیت و تراکم را نشان می‌دهند. آن دسته از شبکه‌های حاصل که بیشترین میزان مرکزیت و تراکم را دارند، ارتباط قوی‌تری بین واژگان کلیدی در بین داده‌های غالب را نشان می‌دهد (کولتر، مونارک، و کوندا، ۱۹۹۸).

در سال‌های گذشته و اخیر، الگوریتم‌ها و روش‌های مختلفی در تحلیل هم‌واژگانی مطرح شده‌اند، مانند شاخص شمول، شاخص مجاورت (کالون، کورتیال، و ترنر، ۱۹۸۶)، شاخص هم‌ارز یا استحکام (کولتر، مونارک، و کوندا، ۱۹۹۸)، شاخص تراکم، شاخص مرکزیت (باین و همکاران، ۱۹۹۱)، بردار راهبردی^۱ (هی، ۱۹۹۹)، تحلیل خوشه‌ای و تحلیل

1. Coulter, Monarch, & Konda

شبکه اجتماعی.

شاخص شمول از محاسبه مفاهیم یا کلیدواژگان کم‌بسامد به دست می‌آید و برای یافتن زمینه‌های مرکزی در یک حوزه پژوهشی به کار رفته است. مقادیر ارزش عددی آن بین صفر و یک است و احتمال مشروط را نشان می‌دهد. یعنی دو واژه در یک مدرک هم‌رخداد هستند؛ در صورتی که تعداد هم‌رخدادی آنها اندک است اما از نظر مفهومی دارای ارتباط هستند. نمایه مجاورت، مفاهیم میانجی را در شبکه نشان می‌دهد. کالون از این نمایه برای شناسایی حوزه‌های کوچک اما مستعد رشد استفاده کرد.

نمایه شمول با فرمول $I_{ij} = \min(C_i, C_j) / C_{ij}$ محاسبه می‌شود. C_{ij} = تعداد مدارکی که کلیدواژه‌های M_i و M_j با هم در آنها وجود دارد. C_j = بسامد رخداد کلیدواژه M_j در مجموعه‌ای از مقالات. C_i = بسامد رخداد کلیدواژه M_i در مجموعه‌ای از مقالات و $\min(C_i, C_j)$ = حداقل دو بسامد C_i و C_j است. شاخص نمایه مجاورت یا نزدیکی با فرمول $P_{ij} = (C_i C_j) / C_{ij}$ محاسبه می‌شود C_i و C_j همان معنای فرمول اول را دارند و N معرف تعداد مقالات مجموعه است.

در مطالعات بعدی، شاخص دیگری معرفی شد که همبستگی ارزش میان دو جفت واژه را محاسبه می‌کرد. کالون و همکاران (۱۹۹۱) این شاخص را شاخص هم‌ارز^۲ و کولتر، مونارک، و کوندا (۱۹۹۸) آن را شاخص استحکام^۳ نامیدند. با توجه به فرمول ۱ فرمول شاخص جدید به این صورت محاسبه شد: $E_{ij} = (C_i / C_j) \cdot (C_j / C_i) \cdot (C_{ij})^2 / (C_i C_j)$ ارزش E_{ij} بین صفر و یک است. به مثابه فرمول اول E_{ij} احتمال حضور همزمان i در مجموعه مدارکی را نشان می‌دهد که با واژه j نمایه شده‌اند و برعکس. به همین دلیل، ترنر و روزان (۱۹۹۱) E_{ij} را "عامل مشترک شمول دوطرفه" نامیدند. پس از محاسبه شاخص شمول و مجاورت، نقشه‌های شمول و مجاورت رسم می‌شود. نقشه‌های شمول موضوعات مرکزی یک حوزه و روابط آنها را با کلیدواژه‌هایی نشان می‌دهد که بسامد رخداد کمتری دارند. نقشه مجاورت، ارتباط میان اندیشه‌های کوچک‌تر یا پنهان در اطراف نقاط مرکزی را نمایش می‌دهد. برای ترسیم نقشه شمول، یالی که بالاترین ارزش را دارد اول انتخاب می‌شود و سپس یال‌های کم‌ارزش تر رسم می‌شوند تا به آستانه I برسیم. کلیدواژه‌هایی که در بالای نقشه قرار می‌گیرند قطب‌های مرکزی یک حوزه پژوهشی را نشان می‌دهند. کلیدواژه‌هایی که هم در موقعیت قطب مرکزی و هم واژه‌های وابسته دارند، واژه‌های "میانجی" نامیده می‌شوند (کالون، کورتیال، و ترنر، ۱۹۸۶؛ عصاره و همکاران، ۱۳۹۴، نقل از نف و کرلی، ۲۰۰۹).

1. Strtegic Diagram
2. Equivalence index
3. Strength index

به دلیل ضعف نمایه‌های شمول و مجاورت در تحلیل هم‌واژگانی سه الگوریتم دیگر پیشنهاد شد. نمودار راهبردی برای شناسایی موضوعات به صورت محلی یا منطقه‌ای و در

سطح جهانی ایجاد شد. در این نمودار موضوعاتی که در سطح جهانی بودند روی محور X قرار گرفت و زمینه‌های علمی که در سطح منطقه‌ای بودند روی محور Y واقع شد. نمودار راهبردی بیشتر برای نشان دادن زمینه‌های پژوهشی در یک فضای دو بُعدی و در قالب چهار ناحیه موضوعی ایجاد شد. از این نمودار تاکنون در تحلیل هم‌واژگانی به‌کار رفته است. نمایه تراکم و نمایه مرکزیت نیز برای محاسبه استحکام و قدرت موضوعات در سطوح منطقه‌ای و جهانی به‌کار گرفته شده است (ژونگ، ۲۰۱۳، نقل از هی، ۱۹۹۹).

نارسایی‌های تحلیل هم‌واژگانی

در تحلیل هم‌واژگانی محدودیت‌هایی وجود دارد که اگر مورد توجه قرار نگیرد تحلیل‌های مورد نظر را با مشکل مواجه می‌کند:

۱. کیفیت واژگان منتخب یکی از مهم‌ترین مراحل تحلیل هم‌واژگانی است. در کیفیت واژگان حوزه مورد نظر، محل استخراج آن در مدرک، غافل شدن از مسائل زبانی واژگان، ترکیب واژگان، ارتباط معنایی واژگان، و تأثیر نمایه‌سازی اهمیت دارد و بی‌توجهی به هر یک از اینها کیفیت واژگان مورد تحلیل را کاهش می‌دهد.

۲. به‌کارگیری این روش در حوزه‌هایی که از لحاظ واژگان و مفاهیم مستعد نباشد از مشکلات اساسی این تحلیل است که بی‌توجهی به آن نتیجه‌گیری پژوهش را دچار تناقض خواهد کرد.

نتیجه‌گیری

تحلیل هم‌واژگانی یکی از روش‌های علم‌سنجی است که در دهه ۱۹۸۰ مطرح شد. این روش بر این فرض استوار است که استفاده از واژه‌های مشترک در دو یا مجموعه‌ای از متون نشان‌دهنده نزدیکی آن متن‌ها به همدیگر است و نیز ابزار قدرتمندی در ردیابی علوم است که از طریق آن می‌توان ساختار، مفاهیم، و مؤلفه‌های یک حوزه علمی را شناسایی، تعیین، و سیاست‌گذاری کرد و پویایی علم و فناوری را به‌تصویر کشید.

تحلیل هم‌واژگانی، با خلاصه‌سازی مدارک در واژه‌هایی قدرتمند و محاسبه رخداد و هم‌رخدادی، تشخیص دقیق‌تری نسبت به حوزه موضوعی ارائه می‌دهد و در کنار سایر تحلیل‌ها همچون تحلیل هم‌استنادی و هم‌نویسندگی، شاخصی مهم در علم‌سنجی است که در مطالعه شبکه مفهومی یا واژگانی یک حوزه، مهم‌ترین حوزه‌های پژوهشی برای شناسایی الگوهای پنهان و برجسته، روابط درونی و بیرونی مفاهیم، رویدادهای در حال ظهور، و سیاست‌گذاری علم و دانش به‌کار گرفته می‌شود. شبکه مفهومی یا واژگانی با روش تحلیل

هم‌واژگانی و با روش هم‌رخدادی واژگان ترسیم و تحلیل می‌شود. این روش به‌عنوان فنی در تحلیل محتوا، کشف روابط پنهان میان مفاهیم و روابط معنایی، و یکی از روش‌های ایجاد روابط سلسله‌مراتبی مفاهیم در هستی‌شناسی حوزه‌های علمی و زمینه‌های دانش تخصصی مطرح است.

تحلیل هم‌واژگانی می‌تواند به ما بگوید که توجه جامعه علمی بیشتر به چه موضوعات، زمینه‌ها، مسائل، پژوهش‌ها، و نظرات پرداخته است؛ حوزه‌ها و زیرحوزه‌های مختلف علمی کدام‌اند و تا چه حد نویسندگان و پژوهشگران در هریک از مقوله‌های علمی مشغول به فعالیت و پژوهش هستند؛ حوزه‌ها و موضوعات مختلف چه سیر تحولی را پشت سر گذاشته‌اند و احتمالاً در آینده چه زمینه‌هایی کانون توجه قرار خواهند گرفت؛ و همچنین چه رابطه مفهومی بین حوزه‌ها و زیرحوزه‌های مفهومی وجود دارد.

مآخذ

- ابویی اردکان، محمد؛ عابدی‌جعفری، حسن؛ و آقازاده، فتاح (۱۳۸۸). کاربرد روش‌های خوشه‌بندی در ترسیم نقشه‌های علم: موردکاوی نقشه علم مدیریت شهری. فصلنامه علوم و فناوری اطلاعات، ۲۵ (۳)، ۳۴۷-۳۷۱.
- احمدی، حمید؛ سلیمی، علی؛ و زنگی‌شینه، الهه (۱۳۹۲). علم‌سنجی، خوشه‌بندی و نقشه دانش تولیدات علمی ادبیات تطبیقی در ایران. کاوش نامه ادبیات تطبیقی، ۳ (۱۱)، ۲۸-۱.
- احمدی، حمید؛ و کویکی، مرتضی (۱۳۹۴). هم‌بندی واژگان: مطالعه‌ای پیرامون پیوند و مرز میان مدیریت اطلاعات و مدیریت دانش براساس انتشارات داخلی نویسندگان ایرانی. پژوهشنامه پردازش و مدیریت اطلاعات، ۳۰ (۳)، ۶۴۷-۶۷۶.
- الهی، شعبان؛ نقی‌زاده، رضا؛ قاضی‌نوری، سپهر؛ و منطقی، منوچهر (۱۳۹۱). شناسایی جریان‌های غالب در حوزه توسعه نوآوری در مناطق با استفاده از روش تحلیل هم‌رخدادی کلمات. فصلنامه بهبود مدیریت، ۱۳۶ (۳)، ۱۵۸-۱۷.
- توکلی‌زاده‌راوری، محمد؛ و نجابتیان، مریم (۱۳۸۹). خوشه‌بندی مبتنی بر مدرک اصطلاح: هم‌جواری موضوعات روان‌شناسی ازدواج در ادبیات زیست پزشکی در دوره‌های زمانی ۱۹۹۹-۱۹۹۰ و ۲۰۰۸-۲۰۰۰. فصلنامه مدیریت اطلاعات سلامت، ۷ (۲)، ۱۷۲-۱۸۶.
- ذوالفقاری، ثریا؛ توکلی‌زاده‌راوری، محمد؛ میرزایی، احمد؛ سهیلی، فرامرز؛ و سجادیان، محمد (۱۳۹۴). کاربرد نقشه‌های حاصل از تحلیل هم‌رخدادی واژگان پروانه‌های ثبت اختراع در آشکارسازی دانش فنی. مطالعات ملی کتابداری و سازماندهی اطلاعات، ۲۷ (۳)، ۱۴۷-۱۵۹.
- سالمی، نجمه؛ و کوشا، کیوان (۱۳۹۲). مقایسه تحلیل هم‌استنادی و تحلیل هم‌واژگانی در ترسیم نقشه کتابشناختی (مطالعه موردی: دانشگاه تهران). پژوهشنامه پردازش و مدیریت اطلاعات، ۲۹ (۱)، ۲۵۳-۲۶۶.

صدیقی، مهتری (۱۳۹۳). بررسی کاربرد روش هم‌رخدادی واژگان در ترسیم ساختار حوزه‌های موضوعی علمی (مطالعه موردی: حوزه اطلاع‌سنجی). پژوهشنامه پردازش و مدیریت اطلاعات، ۳۰ (۲)، ۳۷۳-۳۹۶.
عصاره، فریده؛ سهیلی، فرامرز؛ و منصوری، علی (۱۳۹۴). علم‌سنجی و دیداری‌سازی اطلاعات. اصفهان: انتشارات دانشگاه اصفهان.

کاتبی، فاطمه (۱۹۹۳). تحلیل هم‌رخدادی واژگان حوزه مدیریت دانش در پایگاه وب اوساینده در سال‌های ۱۹۹۳ لغایت ۲۰۱۲ به منظور دیداری‌سازی واژگان تشکیل‌دهنده زیرساخت این حوزه. پایان‌نامه منتشر نشده کارشناسی ارشد، دانشگاه شهید چمران، اهواز.

ناصری‌جزه، محمود؛ طباطبائیان، حبیب‌الله؛ و فاتح‌راد، مهدی (۱۳۹۱). علم‌سنجی و خوشه‌بندی دانش مدیریت فناوری در ایران با هدف ارزیابی این دانش و مقایسه آن با وضعیت جهانی. دومین کنفرانس ملی مدیریت فناوری (تهران، ۲۶ آذر)، بازیابی ۵ اردیبهشت ۱۳۹۳، از http://www.civilica.com/paper-iramot06-iramot06_085.html

Aregon, D. M., & et al. (2005). Domain analysis and information retrieval through the construction of heliocentric maps based on ISI-JCR category cocitation. *Information Processing & Anagement*, 41 (6), 1520-1533.

Bauin, S., Michelet, B., Schweighoffer, M. G., & Vermeulin, P. (1991). Using bibliometrics in strategic analysis: "Understanding chemical reactions" at the CNRS. *Scientometrics*, 22 (1), 113-137.

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. *Scientometrics*, 22 (1), 155-205.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: an introduction to co-word analysis. *Social Science Information*, 22 (2), 191-235.

Callon, M., Courtial, J. P., & Turner, W. A. (1986). Future developments. In M. Callon, J. Law, & A. Rip (Eds.), *Mapping the dynamics of science and technology: Sociology of science in the real world* (pp. 211-217). London: Macmillan.

Cambrosio, A., Limoges, C., Courtial, J. P., & Laville, F. (1993). Historical scientometrics mapping over 70 years of biological safety research with co-word analysis. *Scientometrics*, 27 (2), 119-143.

Cobo, M. J., Lopez-Herrera, A. G., & Herrera-Viedma, E. (2011). An approach for

- detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, 5 (1), 146-166.
- Coulter, N., Monarch, I., & Konda, S. (1998). Software engineering as seen through its research literature: a study in co-word analysis. *Journal of the American Society for Information Science (JASIS)*, 49 (13), 1206-1223.
- Courtial, J. P. (1994). Comments on leydesdorff's a validation study of LEXIMAPP. *Scientometrics*, 49 (1), 98-112.
- De Looze, M. A., & Lemarié, J. (1997). Corpus relevance through co-word analysis: an application to plant proteins. *Scientometrics*, 39 (3), 267-280.
- Ding, Y., Chowdhury, G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using cword analysis. *Information Processing and Management*, 37 (6), 817-842. Retrieved 2014, May 8. Doi:10.1016/S0306-4573(00)00051-0.
- Edwards, D. (1995). Recent advances in descriptive multivariate analysis. *Royal Statistical Society*, 25, 135-156.
- Frantzi, K., Ananiadou, S., Mima, H. (2000). Automatic recognition of multi- word terms". *International Journal of Digital Libraries*, 3 (2), 115-130.
- He, Q. (1999). Knowledge discovery through cword analysis. *Library Trends*, 48 (1), 133-159.
- Hu, C. P., Deng, S.L., & Liu, G. Y. (2013). A co-word analysis of library and information science in China. *Scientometrics*, 97 (1), 369- 382.
- Kaufman, L., & Rousseeuw, P. J. (1990). Finding Groups in Data. *Wiley Series in Probability and Mathematical Statistics*, 123-165. UK: John Wiley and Sons, Inc.
- King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science*, 13 (5), 261-276.
- Kostoff, R. N., Eberhart, H.J, & Toothman, D. R. (1997). Database tomography for information retrieval. *Journal of Information Science*, 23 (4), 301-311
- Leydesdorff, L. (1987). Various methods for the mapping of science. *Scientometrics*, 11 (5-6), 291-320.
- Leydesdorff, L. (1989). Words and co-words as indicators of intellectual organization. *Research Policy*, 18 (4), 209-223.
- Leydesdorff, L. (1991). In search of epistemic networks. *Social Studies of Science*, 21 (1),

- 75-110.
- Leydesdorff, L. (2010). Eugene garfield and algorithmic historiography: Co-words, co-authors, and Journal names. *Annals of Library and Information Studies*, 57(3), 248-260.
- Leydesdorff, L., & Welbers, K. (2011). The semantic mapping of words and co-words in contexts. *Journal of Informetrics*, 5 (3), 417-461.
- Liu, G., Hu, J., & Wang, H. (2012). A Co-word analysis of digital library field in china. *Scientometrics*, 91 (1), 203-217.
- Milojevic, S., Sugimoto, C. R., & Yan, E. (2011). The cognitive structure of library and information science: Analysis of article title words. *Journal of the American Society for Information Science and Technology*; 62 (10), 1933-1953.
- Neff, M. W., & Corley, E. A. (2009). 35 years and 160,000 articles: a bibliometric exploration of the evolution of ecology. *Scientometrics*, 80(3), 657-682.
- Noyons, E., & van Raan, A. (1998). Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. *Journal of the American Society for Information Science*, 49 (1), 68-81.
- Ravikumar S., Agrahari A., & Singh, S. N. (2014). Mapping the intellectual structure of scientometrics: a co-word analysis of the journal. *Scientometrics* (2005-2010); *Scientometrics*, 102 (1), 929-955.
- Rokaya, M. (2008). Ranking of field association terms using co-word analysis. *Information Processing & Management*, 44 (2), 738-755. Doi: 10.1016/j.ipm.2007.06.001
- Rotto, E., & Morgan R. P. (1997). An exploration of expert-based text analysis techniques for assessing industrial relevance in U.S. engineering dissertation abstracts. *Scientometrics*, 40(1), 83-102.
- Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying bibliography and graphing specialties. *Science Studies*, 4 (1), 17-40.
- Turner, W. A., & Rojouan, F. (1991). Evaluating input/output relationships in a regional research network using co-word analysis. *Scientometrics*, 22 (1), 139-154.
- Whittaker, J. (1989). Creativity and conformity in science: Titles, keywords and co-word

- analysis. *Social Studies of Science*, 19 (3), 473-496.
- Yang, Y., Wu, M., & Cui, L. (2012). Integration of three visualization methods based on co-word analysis. *Scientometrics*, 90 (2), 659-673.
- Yin, R., & et al. (2009). *Hotspot for study in UML of China: Co-word analysis*. International Conference on Management of e-Commerce and e-Government, Sept.16-19, (pp. 558-562). Nanchang, China, China. Doi: 10.1109/ICMeCG.2009.92
- Zipf, G. (1949). *Human behavior and the principle of least effort*. Boston: Addison Wesley
- Zong, Q., Shen, H., Yuan, Q., Hu, X., Hou, Z., & Deng, S. (2013). Doctoral dissertations of library and information science in china: a co-word analysis. *Scientometrics*, 94 (2) 781-799.

استناد به این مقاله:

احمدی، حمید؛ عصاره، فریده (۱۳۹۶). مروری بر کارکردهای تحلیل هم‌واژگانی. *مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۲۸ (۱)، ۱۲۵-۱۴۵.