

بررسی مفاهیم، تعاریف و کارکردهای تحلیل هم‌واژگانی^۱

حمید احمدی^۲

فریده عصاره^۳

چکیده:

هدف: در نیم قرن گذشته، ادبیات علم افزایش چشمگیری یافته است، به طوری دست‌یابی به اطلاعات مورد نیاز را با مشکل روبه‌رو کرده است. برخی از فنون و روش‌های کمی برای بهبودی این شرایط توسعه یافته‌اند. روش تحلیل هم‌واژگانی، یکی از این روش‌ها در حوزه علم‌سنجی است.

روش پژوهش: مرور منابع، به روش تحلیلی - سندی

یافته‌ها: یافته‌های این پژوهش شامل چارچوب‌ها، کارکردها، روش‌ها و به‌طور کلی مبانی نظری تحلیل هم‌واژگانی است. به‌طور اخص، شامل پیش‌فرض‌ها موجود، تفاوت‌های آن با سایر روش‌های مشابه، الگوریتم‌ها و فناوری‌های مورد نیاز آن، برای شناخت و استفاده از این روش در تحلیل محتوا می‌باشد.

نتیجه‌گیری: تحلیل هم‌واژگانی یکی از آن روش‌های است که بر اساس هم‌رخدادی واژه‌ها در متون عمل می‌کند و برای کشف روابط و مفاهیم میان افکار، ایده‌های بشری و ردیابی علم در یک زمینه تحقیقاتی مورد استفاده قرار می‌گیرد. طی دو دهه گذشته این روش به‌عنوان یک ابزار قدرتمند در کشف و ترسیم علم و دانش، توسط محققان به کار گرفته شده است.

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

^۱ برگرفته از پایان نامه دکتری دانشگاه شهید چمران اهواز

^۲ دانشجوی دکتری علم اطلاعات و دانش‌شناسی دانشگاه شهید چمران

^۳ استاد گروه علم اطلاعات و دانش‌شناسی و مدیر قطب علمی مدیریت دانش دانشگاه شهید چمران اهواز

مقدمه

در پاسخ به پیشرفت‌های صورت گرفته در زمینه‌ی نگاشت هم‌استنادی در سال‌های دهه‌ی ۱۹۷۰ توسط (اسمال و گریفیث^۱، ۱۹۷۴) افرادی از قبیل (کالون، کورتیال، ترنر و باین^۲، ۱۹۸۳) طرح نگاشت هم‌رخدادی واژگان را به‌عنوان زمینه‌ی تحقیقاتی دیگری برای مطالعه‌ی روابط معنایی در ادبیات علم و فناوری مطرح ساختند (لیدسدورف^۳، ۱۹۸۹، نقل در عصاره، سهیلی و منصور، ۱۳۹۴) تحلیل هم‌واژگانی در کنار سایر روش‌های حوزه علم‌سنجی، می‌توان از طریق آن مدیریت و سیاست‌گذاری علم و دانش و به‌طورکلی تحلیل حوزه را تسهیم کرد.

سنجش و ارزیابی حوزه‌های علمی بدون استفاده از شاخص‌های کمی میسر نیست. این شاخص‌ها که روش معمول تجزیه و تحلیل داده‌ها هستند شامل: زوج کتاب‌شناختی، تحلیل استنادی یا هم‌استنادی، تحلیل هم‌نویسندگی و تحلیل هم‌واژگانی می‌شوند و در حوزه کتاب‌سنجی و علم‌سنجی شناخته شده هستند. در جدول ۱، متداول‌ترین طبقه‌بندی در روش‌های کتاب‌سنجی و علم‌سنجی طبق واحدهای تحلیلی مورد استفاده و روابط ایجاد شده در میان آن‌ها را نمایش می‌دهد (کوبو^۴ و همکاران، ۲۰۱۱).

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
هم‌نویسندگی	نام نویسنده، کشور از وابستگی سازمانی، موسسه از وابستگی سازمانی	هم‌رخدادی نویسندگان، هم‌رخدادی کشورها، هم‌رخدادی مؤسسات
هم‌استنادی	مأخذ نویسنده، مأخذ مجله	نویسنده مورد استناد، مدارک مورد استناد، مجله مورد استناد
هم‌واژگانی	کلمات کلیدی، یا اصطلاحات استخراج شده از عنوان، چکیده، یا مجموعه مدرک	هم‌رخدادی اصطلاحات یا واژه‌ها

کوبو و همکاران، هم‌واژگانی را در کنار سایر روش‌ها، یکی از روش‌های کتاب‌سنجی و علم‌سنجی معرفی می‌کنند. تاکنون در مدارک متعددی به زبان فارسی، روش‌های علم‌سنجی مانند زوج‌های کتاب‌شناختی، تحلیل‌های هم‌استنادی و تحلیل‌های روابط هم‌نویسندگی مورد تبیین قرار گرفته است. در این میان، روش تحلیل هم-

¹ Small & Griffith

² Callon, Courtial, Turner & Bauin

³ Leydesdorff

⁴ Cobo

واژگانی به‌عنوان یک روش در حوزه علم‌سنجی، کمتر از سایر روش‌های علم‌سنجی شناخته شده یا به‌کار گرفته شده است. به نظر می‌رسد تبیین این روش، شناخت مبانی و کارکردهای آن، یک ضرورت باشد. لذا پژوهشگران مقاله حاضر این روش را مسأله مهمی فرض کرده و درصدد هستند آن را مورد تحلیل و بررسی قرار داده و زوایا و کاربردهای آن نیز روشن نمایند. در مقاله حاضر، تعاریف، تاریخچه و پیشینه، کاربردها، الگوریتم‌ها، روش‌شناسی، مزایا و ضعف‌های تحلیل هم‌واژگانی مورد تبیین قرار می‌گیرد.

تعریف‌ها

هم‌واژگانی را از نظر مفهومی معادل واژه Co-word دانسته‌اند که گاهی دیده شده واژه Co-occurrence را به‌جای آن به کار برده‌اند. در زبان فارسی واژه نخست را هم‌واژگانی و واژه دوم را هم‌رخدادی معنا کرده‌اند. تحلیل هم‌واژگانی که بر اساس هم‌رخدادی واژگان عمل می‌کند به‌عنوان یک روش تحلیل محتوا، یکی از روش‌های علم‌سنجی است (کینگ، ۱۹۸۷). تحلیل هم‌واژگانی را به‌عنوان جانشینی برای تحلیل هم‌استنادی در نظر می‌گیرد. (لیدسدورف، ۱۹۹۱) این روش را برای حوزه علم‌سنجی به کاربرد. تحلیل هم‌واژگانی نمونه‌ای از روش مدل‌سازی گرافیکی است که در آن از اندیشه‌های مربوط به تحلیل رابطه استفاده می‌شود (نف و کرلی^۱، ۲۰۰۹). این روش ابزاری قدرتمند در کشف دانش و ترسیم نقشه کتاب‌شناختی است.

علاوه بر تعاریف فوق، تعریف‌های دیگری توسط پژوهشگرانی همچون: (کورتیال، ۱۹۹۴؛ ادواردز^۲، ۱۹۹۵؛ دینگ، چاودری و فو^۳، ۲۰۰۱؛ روکایا^۴، ۲۰۰۸؛ کوستوف^۵، ۱۹۹۸؛ هی،) ارائه شده است که به‌صورت مختصر شامل تعاریف زیر می‌باشد:

- ۱- نوعی تحلیل محتوا برای کشف الگوها و تعیین گرایش‌های موضوعی در حوزه‌های مشخص پژوهشی،
- ۲- اندازه‌گیری درجه ارتباط بین مفاهیم و اصطلاحات،
- ۳- روشی مهم برای دیداری‌سازی روابط میان مفاهیم، اندیشه‌ها و مسائل علمی،
- ۴- تحلیلی کمی به‌منظور کشف شبکه مفاهیم حوزه‌های علمی،
- ۵- روشی برای شناسایی ساختار موضوعی یک حوزه پژوهشی،
- ۶- کشف روابط مفهومی میان مدارک،

¹ Neff & Corley

² Edwards

³ Ding, Chowdhury & Foo

⁴ Rokaya

⁵ Kostoff

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 شده و زمینه را برای بازیابی اطلاعات در پایگاه‌های دانش فراهم می‌آورد.

با توجه به تعارف ذکر شده، تحلیل هم‌واژگانی شیوه‌ای از تحلیل محتوا است که از طریق هم‌رخدادی واژه‌ها یا مفاهیم در متون و مدارک عمل می‌کند و از طریق آن می‌توان مفاهیم اصلی یک زمینه یا حوزه علمی را شناخته و به‌واسطه این شناخت، الگوها و رویدادهای مفهومی حوزه، ترسیم ساختار علمی، شبکه مفهومی، روابط سلسله‌مراتبی مفاهیم و مقولات مفهومی را کشف و در جهت مدیریت حوزه به کار برد. به سخن دیگر، تحلیل هم‌واژگانی ابزاری در جهت کشف الگوهای پنهان و رویدادهای نوظهور مفهومی است.

پیشینه تحلیل هم‌واژگانی

اولین کوشش‌ها در زمینه‌ی واژگان داخل متن توسط (زیف^۱، ۱۹۴۹) صورت گرفت. زیف با مطالعه فراوانی واژه‌هایی که در متون انگلیسی به‌کار رفته‌اند به مصادیقی برای اصل کمترین کوشش دست یافت. تحلیل بسامد واژگان در عناوین مقالات، ابتدا توسط (اسمال و گریفث، ۱۹۷۴) به کار گرفته شد تا موضوع یا به‌طور خاص مفهوم کلی مدرک استناد شده را مشخص کنند.

تحلیل هم‌واژگانی اولین بار در دهه ۱۹۸۰ در فرانسه در مرکز جامعه‌شناسی خلاقیت کول^۲ به‌طور جدی 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 مورد استفاده قرار گرفت.

روش تحلیل هم‌واژگانی در سال‌های اخیر در پژوهش‌های پژوهشگران کشورهای مختلف به نحو قابل ملاحظه‌ای مورد توجه قرار گرفته است. بسیاری از این پژوهشگران با این روش، حوزه‌های خاصی را به قصد تحلیل و ترسیم ساختار آن مورد مطالعه کرده‌اند. نمونه‌هایی از به‌کارگیری این روش در پژوهش‌های (هی، ۱۹۹۹، دینگ و همکاران، ۲۰۰۱، بین^۳ و همکاران، ۲۰۰۹، لیدسدورف، ۲۰۱۰، میلوجویک، سایموتو و یان^۴، ۲۰۱۱، لیو^۵ و همکاران، ۲۰۱۲، یانگ، هو و سیو^۱، ۲۰۱۲، هو، دنگ^۲ و لیو، ۲۰۱۳، زونگ^۳ و همکاران، ۲۰۱۳،

¹ Zipf

² Center De Sociologie de La Innovation of the Ecole

³ Yin

⁴ Milojevic, Sugimoto, & Yan

⁵ Liu

۲۰۱۳، راوی کومار، اگراهاری و سینگ^۱، ۲۰۱۴) دیده می‌شود. در ایران پژوهش مستقلی در مباحث نظری روش تحلیل هم‌واژگانی انجام نگرفته؛ اما با استفاده از کاربردهای این روش، حوزه‌های علمی متعددی مورد تحلیل و مطالعه قرار گرفته است. در این میان می‌توان به پژوهش‌های همچون: (ابوبی اردکانی و همکاران، ۱۳۸۸، توکلی زاده راوری و نجابتیان، ۱۳۸۹، ناصری جزه، طباطبائیان و فاتح راد، ۱۳۹۱، الهی و همکاران، ۱۳۹۱، احمدی، سلیمی و زنگی‌شبه، ۱۳۹۲، سالمی و کوشا، ۱۳۹۲، صدیقی، ۱۳۹۳، کاتبی، ۱۳۹۳، احمدی و، کوبکی، ۱۳۹۴، ذولفقاری و همکاران، ۱۳۹۴) به‌عنوان نمونه اشاره کرد. وجه مشترک این پژوهش‌ها به کاری روش تحلیل هم‌واژگانی برای ترسیم ساختار حوزه‌های مورد مطالعه است. همچنین در بخش روش‌شناسی پژوهش‌های ذکر شده، نکات مفید و فنی درباره این روش بیان شده و وجه مشترک آن، استفاده از بسامد واژگان در تحلیل این روش می‌باشد. همچنین در کل این پژوهش‌ها اشاره مختصری به کاربردهای این روش شده و بیشتر از این فراتر نرفته است.

کاربردهای تحلیل هم‌واژگانی

بررسی مطالعات مختلف در حوزه علم اطلاعات به خصوص در زمینه‌ی تحلیل هم‌واژگانی مانند مطالعات پژوهشگرانی همچون (لاو^۲ و ویتاکر، ۱۹۹۲ و برام و همکاران، ۱۹۹۱) نشان می‌دهد این روش به‌طور کلی می‌تواند در موارد زیر کاربرد داشته باشد:

۱. ترسیم حرکت و پویایی علم (کالون و همکاران ۱۹۸۶)
۲. ترسیم ساختار حوزه‌ها و زمینه‌های علمی (ویتاکر، ۱۳۸۹)
۳. ترسیم روابط میان تحقیقات بنیانی و تحقیقات فناورانه (کالون و همکاران، ۱۳۹۱)
۴. دیداری‌سازی شبکه مفهومی علم و فناوری (سالمی و کوشا ۱۳۹۱)
۵. تحلیل سیر تحول مفهومی در طول دوره‌های زمانی
۶. تعیین و تحلیل نواحی مورد پژوهش

¹ Yang, Wu, & Cui

² Hu, Deng

³ Zong

⁴ Ravikumar, Agrahari & Singh

⁵ Law

۷. ارزیابی درون داد و برون داد روابط در یک شبکه پژوهشی (ترنر و روژان ، ۱۳۹۱، نقل در عصاره و همکاران)
۸. دسته‌بندی مدارک بر اساس موضوعات
۹. خوشه‌بندی مفاهیم حوزه‌های علمی
۱۰. برنامه‌ریزی و سیاست‌گذاری در حوزه‌های علمی
۱۱. درک و کشف روابط پنهان در حوزه‌های علوم
۱۲. کشف الگوهای برجسته و رویدادهای در حال ظهور
۱۳. تعیین روابط سلسله‌مراتبی مفاهیم یک حوزه یا زمینه خاص علمی

پیش‌فرض‌های تحلیل هم‌واژگانی

تحلیل هم‌واژگانی که در سال ۱۹۸۳ توسط کالون مطرح شد، فرض آن این بود که آمدن واژه یا مفاهیم با هم در یک مدرک، نشان‌دهنده محتوای آن مدرک است. لذا اگر میزان این هم‌رخدادی را اندازه‌گیری کنیم، می‌توانیم شبکه مفاهیم یک زمینه علمی را ترسیم کنیم. این شبکه مفهومی با شمارش تعداد دفعات با هم آمدن هر اصطلاح موضوعی با موضوعات دیگر ترسیم می‌شوند 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 هم-رخدادی دو اصطلاح یا دو واژه برای کشف پیوند و رابطه میان دو موضوع در یک حوزه پژوهشی استفاده می‌شود و از این طریق می‌توان توسعه و پیشرفت آن حوزه از علم را پیگیری نمود.

این رویکرد خود بر سه فرضیه استوار است: (۱) کلمات استفاده شده در متون علمی به دقت توسط نویسندگان انتخاب می‌شوند، (۲) استفاده از کلمات مختلف در متن، لازمه وجود برخی روابط غیر جزئی بین آن کلمات است و (۳) تکرار هم‌رخدادی‌های واژه‌ها در متون توسط نویسندگان مختلف، بدین معناست که روابط بین این واژه‌ها، در حوزه علوم که مورد مطالعه واقع می‌شوند، اهمیت دارد (میلوجویک، ۲۰۰۹ نقل در عصاره و همکاران، ۱۳۹۴).

ویتاگر (۱۹۸۹) و هی (۱۹۹۹) در پژوهش‌های خود پیش‌فرض‌های دیگری برای این تحلیل ذکر می‌کنند:

- پدیدآورندگان در انتخاب اصطلاحات موضوعی نهایت دقت دارند.
- به کاری واژه‌ها در مدارک، نشان‌دهنده ارتباط مستقیم آن با محتوای مدارک است
- واژگان داخل متن نشان‌دهنده روابط معنایی موضوع یا حوزه مورد نظر است
- کلیدواژه‌های توصیفی نمایه‌سازان منبع خوبی در تحلیل هم‌واژگانی می‌باشد.

تفاوت تحلیل هم‌واژگانی با تحلیل هم‌استنادی^۱

تفاوت مهمی که میان تحلیل هم‌واژگانی و هم‌استنادی وجود دارد این است که تحلیل هم‌استنادی در دوره‌های پژوهشی مشخص به منابع استناد دهنده و مأخذ استناد شده 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 هم‌استنادی در خوشه‌های سلسله مراتبی یا مقیاس چندبعدی به نمایش گذاشت.

تحلیل هم‌واژگانی را ابزار مناسب‌تری برای مطالعه روابط درونی میان پژوهش‌های دانشگاهی و فناورانه دانسته‌اند، زیرا شاخص‌های تحلیل هم‌استنادی تنها وجود ارتباط را نشان می‌دهد و اطلاعاتی از موضوع مورد پرسش را ارائه نمی‌دهند. برای درک این مسأله که آیا تحقیقات علمی عامل اصلی اختراع و ابداع است یا فناوری، باید به خود متون مراجعه کرد و محتوای مدارک و اختراعات را بررسی کرد که این کار با تحلیل واژگان داخل متون فراهم می‌گردد.

منابع اطلاعاتی در تحلیل هم‌واژگانی

بررسی متون نشان می‌دهد که دامنه واژه‌های مورد استفاده در تحلیل هم‌واژگانی، از کلیدواژه‌های یک اصطلاح‌نامه تا واژه‌های داخل یک متن کامل می‌تواند گسترده باشد. اولین مطالعات تحلیل هم‌واژگانی بر اساس کلیدواژه‌های اصطلاح‌نامه‌ای انجام شده است (باین، ۱۹۸۶ نقل در هی، ۱۹۹۹). بعد از آن متون بر اساس عنوان، خلاصه، و یا تعداد مشخصی از کلیدواژه‌های منحصر به توصیفگرهای یک اصطلاح‌نامه مورد مطالعه قرار گرفتند (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 تمام متن، توانایی بازیابی عبارات مهم و در عین حال کم بسامد است، که در سایر تحلیل‌ها نادیده گرفته می‌شود.

نویون و ون‌ران^۴، ۱۹۹۸ این فرآیند انتخاب واژگان را نشانه DNA یک مدرک قلمداد کرده‌اند. در تمام انتشارات، کلیدواژه‌ها آگاهانه انتخاب می‌شوند و عبارتی را تشکیل می‌دهند تا صفات یک رشته را توصیف کنند. به سخن دیگر، مجموعه هزاران مدرک یک حوزه پژوهشی شبکه عظیمی را تشکیل می‌دهند که در آن، کلیدواژه‌ها، مدارک را به هم پیوند می‌دهند. هرچه واژه‌های مشترک دو مقاله بیشتر باشد، این مقالات ارتباط

¹ Co- Citation Analysis

² Roto & Morgan

³ Eberhart & Toothman

⁴ Noyons, & van Raan

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. فنون ریاضی برای بازنمایی این شبکه‌های انتشاراتی مورد استفاده قرار گرفته‌اند. شباهت واژگانی خوشه‌بندی و نقشه دو وجهی از خوشه‌ها، ابزارهایی است که این فرآیند ساختار پنهان را به تصویر می‌کشد. این ساختارها را می‌توان ساختار شناختی حوزه علمی دانست.

کیفیت نتایج حاصل از تحلیل هم‌واژگانی به عوامل زیادی مانند کیفیت کلیدواژه‌ها و واژه‌های نمایه‌ای، دامنه پایگاه اطلاعاتی، قابلیت روش‌های آماری و ارائه یافته‌ها بستگی دارد. مهم‌ترین مسأله در این روش، انتخاب کلیدواژه است که از آن به «تأثیر نمایه‌ساز» یاد شده است و اغلب محققان بر آن تأکید کرده‌اند. نمایه‌سازی مداخله تحلیل‌گر با متن است و اعتبار نقشه، بستگی به ماهیت نمایه‌سازی دارد. ویتاکر (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 حائز اهمیت است. از آنجایی که منابع مورد نیاز برای این تحلیل، واژگان یا مفاهیم هستند، اگر انتخاب زمینه و حوزه‌ای که دارای مفاهیم غنی و یا از لحاظ جامعیت و مانعیت مفهومی قابل طرح باشد نتایج دقیق‌تری حاصل می‌شود. بنابراین انتخاب حوزه یا زمینه مورد مطالعه با این روش، بایستی با محدودیت‌ها و کارکردهای این روش هم‌خوان باشد.

گام دوم، استخراج مفاهیم یا واژگان از مدارک

دو شیوه برای استخراج مفاهیم از مدارک 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، ۱۹۸۸) 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 کرده و توصیفگرها را قابل اعتمادتر از واژه‌های نمایه‌ساز می‌داند. (ویتاگر، ۱۹۸۹) بر خلاف 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 واژه-ها، 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، سامان‌دهی کرد. ساختار معنایی در روابط بین واژگان ممکن است در این رده‌های مختلف متفاوت باشد (لیدسدورف و ولبرز^۱، ۲۰۱۱).

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

شیوه دوم، استخراج مفاهیم به‌طور مستقیم از متن مدارک است. این شیوه به دلیل گسترده بودن متن‌ها، کاری سخت و زمان‌بر است و هزینه گردآوری آن هم بالاست. لذا در این شیوه استخراج، معمولاً با روش متن-کاوی به‌صورت خودکار انجام می‌گیرد.

گام سوم، آماده‌سازی و انتخاب نهایی مفاهیم

مرحله آماده‌سازی و انتخاب نهایی واژگان یا مفاهیم از اهمیت ویژه‌ای برخوردار است و به خاطر مواردی که مطرح می‌شود، چنانچه با دقت انجام نگیرد در مراحل بعدی و در نتیجه‌گیری، پژوهش دچار نقص خواهد شد. در این مرحله معمولاً چهار عمل صورت می‌گیرد که شامل: یکدست‌سازی واژه‌ها، برچسب‌گذاری، سیاهه

¹ Welbers

² Cambrosio & Limoges

³ Coulter, Monarch & Konda

⁴ De Looze, & Lemarié

کلمات بازدارنده و انتخاب روش و عملی که به وسیله آن از میان واژه‌های استخراج شده اولیه، مفاهیم و یا کلیدواژگان با اهمیت انتخاب شود. اجرای کارهای فوق آسان نخواهد بود. زیرا هر یک از آن‌ها تابع قواعد خاصی هستند و بایستی با دقت انجام گیرند. مثلاً استفاده از پالایه‌های زبانشناسی در ترکیب‌بندی واژگان، مسائل دستوری، حروف اضافه، به کارگیری فرم جمع یا مفرد واژگان و غیره.

بنابراین محققان باید ابتدا تعیین کنند که چه چیزی را به عنوان واحدهای مرتبط با تحلیل در نظر می‌گیرند. دوم اینکه چه کلماتی را باید در تحلیل وارد ساخت؟ گزینه‌های مشخصی که تاکنون مورد استفاده قرار گرفته است. در این مقاله به دو روشی که باعث انتخاب و استخراج مفاهیم با اهمیت می‌گردد، پرداخته می‌شود. یکی از آن‌ها استفاده از روش **C-value** و دیگری عمل وزن دهی به مفاهیم است. روش **C-value** از جمله روش‌هایی است که ترکیبی از روش‌های زبانی و آماری را برای استخراج واژه‌های ترکیبی استفاده می‌کند (فرانتزی، آنانیودیو و می-ما، ۲۰۰۰). روش زبانی **C-value** شامل برچسب‌گذاری بخشی از کلام، پالایه‌ی زبانی و سیاهه‌ی کلمات توقف است. برچسب‌گذاری بخشی از کلام، اختصاص دادن برچسب‌های گرامری (نظیر اسم، صفت، فعل، حرف اضافه، ضمیر، قید) به هر واژه در متن است. روش آماری **C-value**، روشی برای استخراج واژه‌های چند کلمه‌ای است که هدف آن بهبود استخراج واژه‌های طولانی‌تر است که ممکن است به تنهایی در متن رخ ندهند. پالایه زبانی: اسم + اسم و صفت + اسم بهترین روش برای انتخاب کلیدواژگان با کیفیت است. به منظور محاسبه‌ی مقدار **C-value** رشته‌ی **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 بر اساس رابطه‌ی زیر به دست می‌آید که در آن $|a|$ طول رشته‌ی **a** و $f(a)$ فراوانی رخداد آن در متن است: (۲)

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

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 به عنوان یک رشته تودرتو و تعداد واژه‌های طولانی‌تر محاسبه شود. در این حالت مقدار **C-value** بر اساس رابطه‌ی ۳ محاسبه می‌شود:

$$C - value(a) = \log_2 |a| \cdot (f(a) - \frac{1}{P(T_a)} \sum_{b \in T_a} f(b)) \quad (3)$$

که در آن $|a|$ طول رشته‌ی **a**، $f(a)$ فراوانی رخداد رشته‌ی **a** در متن، T_a مجموعه‌ی رشته‌های نامزد استخراج شده شامل **a**، $P(T_a)$ تعداد عناصر T_a و $\sum_{b \in T_a} f(b)$ مجموع فراوانی‌هایی است که **a** در رشته‌های طولانی‌تر رخ می‌دهد (همان). به منظور درک بهتر نحوه‌ی محاسبه، به نمونه زیر اشاره خواهد شد.

¹ Frantzi, Ananiadou & Mima

چنانچه دو مفهوم مانند «ضریب تأثیر» با تعداد رخداد ۲۰ و «ضریب تأثیر مجلات» با رخداد ۸ داشته باشیم. ارزش C-value آن‌ها بر اساس محاسبه فرمول فوق به ترتیب ۱۱/۰۶ و ۱۶/۵ است:

$$\log_2|3|.8 = 11.06 \quad \text{و} \quad \log_2|2|. (20 - \frac{1}{2}(8)) = 16/5$$

باین حال، سالتون و مک گیل^۱ (۱۹۸۳) معتقدند که واژه‌های با بیشترین و کمترین فراوانی ممکن است کم اهمیت‌تر از واژه‌های با فراوانی متوسط باشند. بدین منظور، این نویسندگان سنجه‌ای را با عنوان روش معمول TF-IDF «فراوانی واژه-معکوس فراوانی سند» مطرح کردند؛

TF بیانگر نسبت مدارک دربرگیرنده آن مفهوم در بین تمامی مدارک است. برای محاسبه‌ی 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 یعنی رابطه:

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

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

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

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

این رابطه نشان دهنده‌ی حاصل ضرب TF در IDF و بیانگر اهمیت یک واژه

یا مفهوم در مدرک بوده و می‌توان بر اساس آن مفاهیم موجود در مدارک را بر حسب میزان اهمیت آن‌ها رتبه‌بندی کرد برای درک بیشتر، چنانچه مفهوم «اخلاق علمی» با بسامد ۴۲ در یک حوزه‌ای که یکی از مفاهیم آن با بالاترین (۶۰) بسامد آمده باشد، وزن مفهوم «اخلاق علمی» نسبت به کل حوزه به شرح زیر محاسبه می‌گردد. در این حوزه فرضی ۲۰ مدرک در مورد 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 ۱/۵۷، مطابق فرمول سوم $1/0.9 = 1/0.7 \times 1/0.57$. بدین صورت وزن مفهوم اخلاق علمی برابر تقریباً ۱ است. بدین ترتیب با انتخاب یک آستانه مورد نظر می‌توان با انجام محاسبات فوق در یکی از دو روش مذکور، مفاهیم یا واژگان با اهمیت را در تحلیل هم-واژگان انتخاب کرد و نتیجه بهتری حاصل کرد. علاوه بر آن، معمولاً از روش وزن دهی برای تعیین خوش‌های مفهومی در تحلیل هم‌واژگانی مورد استفاده قرار می‌گیرد.

¹ Salton & McGill

به‌طور خلاصه پژوهشگر می‌تواند از چهار شاخص برای انتخاب سیاهه واژگان دخیل در تحلیل استفاده کند: (۱) فراوانی واژه، (۲) مقدار TF-IDf، (۳) مشارکت ستون‌ها با ماتریس خبی دو و (۴) مجموع تفاضل مقدار مشاهده شده / مقدار پیش بینی شده برای هر کلمه. مطالعه‌ای که توسط لیدسدورف و ولبرز (۲۰۱۱) صورت گرفت، استفاده از سنج‌های اخیر راحت‌تر بوده است. هرچند که هر چهار روش قابل سنجش هستند.

مرحله چهارم: ساخت ماتریس هم‌رخدادی واژگان

تحلیل ویژگی‌ها و شاخص‌های ماتریس هم‌رخدادی مهم‌ترین مرحله از تحلیل هم‌رخدادی واژگان است. در اغلب موارد دو هدف مورد نظر است، یکی شناسایی سلسله مراتبی مفاهیم در میان حوزه‌های یک مسأله پژوهشی، دوم شناسایی حوزه‌های کوچک اما مستعد رشد. کالون برای پاسخ به این دو درخواست، دو شاخص معرفی می‌کند: شاخص شمول^۱ و شاخص مجاورت^۲ (کالون و همکاران، ۱۹۸۶). ماتریس پایه در تحلیل، 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 ماتریس واژه-واژه و از ماتریس واژه-سند برای تعیین روابط سلسله مراتبی مفاهیم استفاده به عمل می‌آید.

ارزش هر سلول در ماتریس بر اساس تعداد دفعاتی که در دو واژه یا مفهوم در کنار یکدیگر در یک مدرک ظاهر می‌شوند، تعیین می‌شود. در این حالت بالاترین میزان رخداد دو واژه در ماتریس به معنای نزدیک‌ترین ارتباط بین دو واژه می‌باشد (روکویا^۳، ۲۰۰۸). ماتریس واژه-واژه از نوع متقارن است یعنی سطر و ستون‌ها برابر هستند. اما در ماتریس واژه-سند می‌تواند نامتقارن باشد و ستون‌ها سند یا مدارک و سطرها مفاهیم این ماتریس را تشکیل می‌دهد. به عبارتی در این ماتریس متون برای بیان حالتی (مثل ردیف‌ها) در نظر گرفته می‌شوند که در آن واژگان به‌عنوان متغیر (ستون‌ها) به آن‌ها نسبت داده می‌شود. (لیدسدورف و ولبرز، ۲۰۱۱). این دو ماتریس هر کدام با هدف‌های خاصی ساخته می‌شود. معمولاً برای ایجاد شبکه مفهومی یا واژگانی از ماتریس متقارن و برای استخراج روابط سلسله مراتبی مفاهیم از ماتریس نامتقارن استفاده می‌شود.

گام پنجم: تحلیل و ترسیم نقشه‌های مفهومی با داده‌های هم‌رخدادی

¹ Inclusion index

² Proximity index

³ Rokaya

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 می‌گردد. یک 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 پژوهش و دستیابی به مزایای بلند مدت 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 یک رشته علمی، راه را برای شناسایی هرچه بهتر و دقیق‌تر آن، شاخه از دانش بشری و تبدیل مفهوم انتزاعی رشته علمی به مفهومی عینی‌تر هموار کردن (آنگون^۱ و همکاران ۲۰۰۵). این نقشه با فنون و روش‌های متعددی ترسیم می‌شود. همان‌گونه که پیش‌تر گفته شد یکی از اهداف و کاربردهای تحلیل هم‌واژگانی، ترسیم ساختار علم یا همان ترسیم نقشه‌های علمی است. کیفیت نقشه‌ها به تعداد مدارک، تعداد مفاهیم یا کلیدواژه، نوع ماتریس و نرم‌افزار، بستگی دارد. تحلیل نقشه‌های ترسیم شده نیاز به مهارت فنی و آگاهی از حوزه مورد نظر دارد.

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 راحت‌تر شده و با دقت زیادی انجام می‌گیرد. در سال‌های اخیر نرم‌افزارهای زیادی در ترسیم و تحلیل داده‌های شبکه‌ای به خصوص در تحلیل هم‌واژگانی به کار گرفته شده است. نمونه این نرم‌افزارها شامل : Biblexcel ، Ucinet ، Pajek ، CiteSpaceII ، Sci Tool ، Network Workbench ، Fulltext.exe ، NetDraw ، VOSViewer ، IN-SPIRE ، ColPalRed ، Google Maps ، Google Earth ، nodexl . است. 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 راه استفاده می‌کند. راه اول، 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 می‌گردند، پیوند بیرونی نام دارند (کافمن و روسو، ۱۹۹۰). دو واژه کلیدی را همایند یا هم‌واژگانی گویند که در متن‌های ثبت شده و در پایگاه داده با هم نمود یافته باشند. فرض کنید C_k تعداد رخدادهای واژه کلیدی k در کل متون یک مجموعه داده باشد. اگر C_{ij}

¹ Aneğón

تعداد رخدادهای واژگان کلیدی i و j باشد، میزان پیوستگی و ارتباط بین دو توصیفگر i و j از طریق رابطه‌ی زیر به دست می‌آید: (S : عدد صحیح \times عدد صحیح)

تابع حقیقی $(c_{ij}, c_i, c_j) : S = \text{اگر } ((c_i \leq 0) \vee (c_j \leq 0))$ آنگاه S = تعریف نشده؛ در 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؛ این شبکه قوی‌ترین ارتباطات موجود در مجموعه داده‌ها را به تصویر می‌کشد. عددی که نشان دهنده‌ی 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 مسأله 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 کمینه باشد (کولتر و همکاران، ۱۹۹۸)

در روش اول، 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 به کار نرفته، به گرافی پیوند می‌زند که در حال شکل‌گیری است و قبل از همه به شبکه اضافه شده است.

شبکه‌ی بعدی به روش مشابهی تولید می‌شود؛ شکل‌گیری آن با پیوند با بزرگ‌ترین عدد هم‌آیندی که در گراف موجود نیست، آغاز می‌شود. در روش دوم الگوریتم، با انتخاب پیوندهایی که از بزرگ‌ترین میزان هم‌رخدادی برخوردارند و از آستانه‌ی هم‌واژگانی فراتر هستند و در برخی از شبکه‌های روش یک وجود دارند، به هر یک از گراف‌های موجود گره‌هایی اضافه می‌شود. شبکه‌های تولید شده 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 قوی‌تر بین واژگان کلیدی در بین داده‌های غالب را نشان می‌دهد (همان)

در سال‌های گذشته و اخیر الگوریتم‌ها و روش‌های مختلفی در تحلیل هم‌واژگانی مطرح شده‌اند. از جمله این روش‌ها و الگوریتم‌ها می‌توان به شاخص شمول، شاخص مجاورت (کالون، ۱۹۸۶)، شاخص هم‌ارز یا استحکام (کولتر و همکاران، ۱۹۹۸)، شاخص تراکم، شاخص مرکزیت (باین و همکاران، ۱۹۹۱)، بردار راهبردی^۱ (هی، ۱۹۹۹)، تحلیل خوشه‌ای و تحلیل شبکه اجتماعی اشاره کرد.

¹ Strategic Diagram

شاخص شمول از محاسبه مفاهیم یا کلیدواژگان کم بسامد 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 کم است. اما از نظر مفهومی دارای ارتباط هستند. نمایه مجاورت، همان گونه از معنای آن پیداست مفاهیم میانجی را در شبکه نشان می‌دهد. کالون از این نمایه برای شناسایی حوزه‌های کوچک اما مستعد رشد استفاده کرد.

نمایه شمول با فرمول: $I_{ij} = C_{ij} / \min(C_i, C_j)$ محاسبه می‌شود که در آن C_{ij} = تعداد مدارکی که کلیدواژه-های M_i و M_j با هم در آن‌ها وجود دارد، می‌باشد. C_j = بسامد رخداد کلیدواژه M_j در مجموعه‌ای از مقالات است. C_i = بسامد رخداد کلیدواژه M_i در مجموعه‌ای از مقالات و $\min(C_i, C_j)$ = حداقل دو بسامد C_j و C_i می‌باشد. شاخص نمایه مجاورت یا نزدیکی با فرمول $P_{ij} = (C_{ij} / C_i C_j) \cdot N$ محاسبه می‌گردد و در این فرمول C_i و C_j و C_{ij} 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^۱ و کولتر و همکارانش (۱۹۹۸) آن را شاخص استحکام^۲ نامیدند. با توجه به فرمول ۱، فرمول شاخص جدید به این صورت محاسبه می‌شد: $E_{ij} = (C_{ij} / C_i)$. $E_{ij} = (C_{ij} / C_j) = (C_{ij})^2 / (C_i \cdot C_j)$ ارزش E_{ij} بین صفر و یک است. به مثابه فرمول اول E_{ij} احتمال حضور همزمان i در مجموعه مدارکی که با واژه j نمایه شده‌اند را نشان می‌دهد و برعکس. به همین دلیل (ترنر و روژان، 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 که بسامد رخداد کمتری دارند را نشان می‌دهد. 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 می‌شد و سپس یال‌های کم‌ارزش‌تر 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 و همکاران، ۱۹۸۶؛ نف و کرلی، ۲۰۰۹ نقل در عصاره و همکاران).

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

¹ Equivalence Index

² Strength Index

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

برخی از نارسایی‌های تحلیل هم‌واژگانی

در تحلیل هم‌واژگانی محدودیت‌هایی وجود دارد که چنانچه مورد توجه قرار نگیرد تحلیل‌های مورد نظر را دچار مشکل خواهد کرد. برخی این محدودیت‌های این روش شامل

۱. کیفیت واژگان: 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 عدم توجه به هر یک از این موارد باعث عدم کیفیت واژگان مورد تحلیل خواهد شد.

۲. به کارگیری این روش در حوزه‌هایی که از لحاظ واژگان و مفاهیم مستعد نباشد، از جمله مشکلاتی اساسی در این تحلیل است که عدم توجه به آن نتیجه‌گیری پژوهش را دچار تناقض خواهد کرد.

نتیجه‌گیری

تحلیل هم‌واژگانی یکی از روش‌های علم‌سنجی است که در دهه ۸۰ میلادی مطرح شد. این روش بر این فرض استوار است که چنانچه در دو یا مجموعه‌ای از متون، واژه‌های مشترکی استفاده شده باشد نشان دهنده نزدیکی آن متن‌ها به همدیگر می‌باشد. همچنین ابزار قدرتمندی در ردیابی علوم است و از طریق آن می‌توان ساختار، مفاهیم و مؤلفه‌های یک حوزه علمی را شناسایی، تعیین و سیاست‌گذاری کرد. همچنین از طریق آن می‌توان پویایی علم و فناوری را ترسیم کرد.

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

در کشف روابط پنهان میان مفاهیم و روابط معنایی و یکی از روش‌های ایجاد روابط سلسله مراتبی مفاهیم در هستی‌شناسی حوزه‌های علمی و زمینه‌های دانش تخصصی می‌باشد.

تحلیل هم‌واژگانی قادر است به ما بگوید که توجهات جامعه علمی، بیشتر به چه موضوعات، زمینه‌ها، مسائل، پژوهش‌ها و نظرات پرداخته است؛ حوزه‌ها و زیر حوزه‌های مختلف علمی کدامند و تا چه حد، نویسندگان و پژوهشگران در هر یک 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 حوزه‌های و زیر حوزه‌های مفهومی وجود دارد

فهرست منابع

- ابویی اردکان، محمد، عابدی جعفری، حسن، و آقا زاده، فتاح (۱۳۸۸). کاربرد روش‌های خوشه‌بندی در ترسیم نقشه‌های علم: مورد کاوی نقشه علم مدیریت شهری. فصلنامه علوم و فناوری اطلاعات، ۲۵ (۳)، ۳۴۷-۳۷۱.
- احمدی، حمید، سلیمی، علی، و زنگی‌شاه، الهه (۱۳۹۲). علم‌سنجی، خوشه‌بندی و نقشه دانش تولیدات علمی ادبیات تطبیقی در ایران. کاوش نامه ادبیات تطبیقی، ۳ (۱۱)، ۱-۲۸.
- احمدی، حمید، و کوبی، مرتضی (۱۳۹۴). هم‌بندی واژگان: مطالعه‌ای پیرامون پیوند و مرز میان مدیریت اطلاعات و مدیریت دانش بر اساس انتشارات داخلی نویسندگان ایرانی. پژوهش‌نامه پردازش و مدیریت اطلاعات، ۳۰ (۳)، ۶۴۷-۶۷۶.
- الهی، شعبان، نقی‌زاده، رضا، قاضی نوری، 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)، ۱۷-۱۵۸.
- توکلی‌زاده راوری، 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، توکلی‌زاده راوری، محمد، میرزایی، احمد، سهیلی، فرامرز، سجادیان، محمد (۱۳۹۴). بررسی کاربرد نقشه‌های حاصل از تحلیل هم‌رخدادی واژگان پروانه‌های ثبت اختراع در آشکارسازی دانش فنی، فصلنامه مطالعات ملی کتابداری و سازمان‌دهی اطلاعات (زود آیند)
- ذوالفقاری ثریا، توکلی زاده راوری، محمد، میرزایی، احمد؛ سهیلی، فرامرز (۱۳۹۴). آشکارسازی زمینه‌های موضوعی فناوری برق در ربات‌های زیرسطحی خودکار از طریق تحلیل هم‌واژگانی پروانه‌های ثبت اختراع فصلنامه رهیافت، ۵۹ (۵۹)، ۵۱-۶۴.
- 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، مه‌ری. (۱۳۹۳). بررسی کاربرد روش هم‌رخدادی واژگان در ترسیم ساختار حوزه‌های موضوعی علمی (مطالعه موردی: حوزه اطلاع‌سنجی). مجله پردازش و مدیریت اطلاعات، ۳۰ (۲)، ۳۷۳-۳۹۶.

- عصاره، فریده، سهیلی، فرامرزی، و منصور، علی (۱۳۹۴). علم‌سنجی و دیداری‌سازی اطلاعات. اصفهان: انتشارات دانشگاه اصفهان کاتبی، فاطمه (۱۹۹۳). تحلیل هم‌رخدادی واژگان حوزه‌ی مدیریت دانش در پایگاه وب او سایند در سال‌های ۱۹۹۳ لغایت ۲۰۱۲ به‌منظور دیداری‌سازی واژگان تشکیل دهنده‌ی زیرساخت این حوزه. پایان‌نامه منتشرنشده کارشناسی ارشد، دانشگاه شهید چمران. دانشکده علوم تربیتی و روانشناسی. گروه علم اطلاعات و دانش‌شناسی.
- ناصری جزه، محمود، طباطبائی، حبیب‌الله، و فاتح راد، مهدی (۱۳۹۱). علم‌سنجی و خوشه‌بندی دانش مدیریت فناوری در ایران با هدف ارزیابی این دانش و مقایسه آن با وضعیت جهانی. مقاله منتشر شده در کنفرانس ملی مدیریت فناوری (تهران)، بازبایی شده در ۵ اردیبهشت ۱۳۹۳ از http://www.civilica.com/paper-iramot06-iramot06_085.html
- Baain, 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.
- Boyack K., Wylie B. N., & Davidson G.S. (2002). Domain visualization using VxInsight for science and technology management. *Journal of the American Society for Information Science and Technology*, 53 (9), 764-774.
- Braam, R., Moed, H. F. & van Raan A. F. (1991). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for Information Science*, 42 (4): 233-251.
- 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., & Baain, 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* London: Macmillan. 211-217.
- 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, 817-842. Retrieved 2014, May 8, from: 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.
- 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
- Law, J., & Whittaker, j. (1992). Mapping acidification research: a tes of the co-word ethod. *Scentometrics* 23 (3): 417-461.
- Leydesdorff, L. (1987). Various methods for the Mapping of Science. *Scientometrics* 11: 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*. NewYork: 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, ICMecG 2009* , art. no. 5279781 , pp. 558-562. Retrieved 2013 October 20, from: . Toc.proceedings.com/06360webtoc.pdf.
- 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.