نقشه‌های علمی: فنون و روش‌ها

هادی رمضانی[۱]
مهدی علیپور حافظی[۲]
عصمت مؤمنی[۳]

چکیده

با رشد حوزه‌های دانش، انتشارات علمی نیز به سرعت توسعه یافته و در نتیجه این رشد، رصد روندهای پژوهشی در حوزه‌های گوناگون علم دشوار شده است. با توجه به محدودیت‌های روش‌های مرسوم تجزیه و تحلیل حوزه‌های مختلف علوم در دنیای جدید و فقدان تصور کامل و جامع آنها از علوم مختلف، روش‌های نوینی توسط متخصصان علم‌سنجی و علوم رایانه ارائه شده است. در این راستا، متخصصان علم‌سنجی در تلاشند تا نقشه‌های حوزه‌های مختلف علمی را ترسیم کنند. به همین دلیل، آنها شاخص‌ها و روش‌های علم‌سنجی را با فنون نوین مصورسازی ترکیب کرده‌اند تا برای سیاست‌گذاران و جوامع علمی، بررسی و پیمایش در مسیرهای مختلف آن میسر شود.

مطالعه حاضر، با هدف معرفی روش‌های نگاشت نقشه‌های علمی؛ به خصوص تحلیل هم‌رخدادی واژگان و روش‌های متن‌کاوی، سعی دارد تا نظر پژوهشگران را به سمت حوزه‌های جدید تحلیل ساختار علوم رهنمون سازد. این مقاله ضمن توجه بر مفاهیم نگاشت نقشه‌های علمی و بررسی روش‌های آن، به توضیح بنیان‌های نظری نقشه‌های علمی می‌پردازد و در نتیجه این نوع نقشه‌ها را به عنوان راه‌حلی برای بازنمون دانش ضمنی در راستای تبدیل آن به دانش آشکار، برای جامعه علمی پیشنهاد می‌کند.

کلیدواژه‌ها: علم‌سنجی، نقشه‌های علمی، مصورسازی، داده‌کاوی، متن‌کاوی

مقدمه

یکی از نخستین ملزومات علاقه‌مندان و پژوهشگران برای ورود به هر حوزه علمی، داشتن درکی صحیح از آن علم است؛ بنابراین، معرفی مباحثی که در محدوده هر علم می‌گنجد بیش از هر چیز، ضروری محسوب می‌شود و می‌تواند زمینه‌های آشناییِ‌شان را فراهم کند. در این میان، بیان مفهوم،

۱. دانشجوی کارشناسی‌ارشد علم اطلاعات و دانش‌شناسی، دانشگاه علامه طباطبائی؛ hadiramazani14@gmail.com
۲. استادیار علم اطلاعات و دانش‌شناسی، پژوهشگاه علوم و فناوری اطلاعات ایران؛ alipour@irandoc.ac.ir
۳. استادیار علم اطلاعات و دانش‌شناسی، دانشگاه علامه طباطبائی؛ momeni.esmat@yahoo.com

سابقه، چهارچوب، دامنه، اجزاء و کارکردهای هر علم و همچنین تحلیل و بررسی جایگاه و روابط آن در زنجیره به هم تنیده علوم بشری به ویژه بیان این روابط با حوزه‌هایی که با آن وابستگی بیشتری دارند، دارای جایگاه برجسته است. در مجموع، این آشنایی باید شرایطی را فراهم آورد که در نهایت، تصویری درست از زمینه‌های فعالیت و کاربرد آن علم ارائه شود و این تصویر، راهگشای کسانی باشد که هنوز مسیرهای پژوهشی آینده خود را تعیین نکرده‌اند (نوروزی چاکلی، ۱۳۹۰).

در این راستا، در دهه‌های ۱۹۵۰ و ۱۹۶۰ به دنبال افزایش میزان اطلاعات و گسترش تولیدات علمی و رواج روش‌شناسی پوزیتیویسم[1] رویکردهای کمّی برای سنجیدن میزان تولید اطلاعات علمی در حیطه‌های گوناگون علم وارد (عابدی جعفری و همکاران، ۱۳۹۰) و بر این اساس روشهای متعددی برای ارزیابی و سنجش تولیدات علمی طراحی و ایجاد شدند که «علم‌سنجی»[2] یکی از متداول‌ترین روش‌های ارزیابی فعالیت‌های علمی و مدیریت پژوهش است. در جریان این خط سیر بررسی کمّی تولیدات علمی، سیاست‌گذاری‌های علمی و فنّاورانه (شامل روش‌ها و حوزه‌های علم‌سنجی)، ارتباطات علمی دانش‌پژوهان، طرح نقشه‌های معرفت‌شناختی و علمی حوزه‌های مختلف دانش، و غیره، برخی از موضوع‌های حوزه علم‌سنجی هستند. همچنین مطالعات علم‌سنجی در جستجوی پاسخ به این سؤال اساسی است که «تحولات علمی چگونه صورت می‌گیرد و مسیری که هر رشته علمی طی می‌کند، به چه صورت است؟» (عصاره و همکاران، ۱۳۸۸). علم‌سنجی با بررسی و کشف نظام و ساختار یک حوزه علمی به روش‌های کمّی، دستاوردهای یک قلمرو فکری را معین و حتی خطوط احتمالی برای پیشرفت‌های بعدی را پیش‌بینی می‌کند. هر چند علم و پژوهش‌های علمی، فعالیتی چند بُعدی است که باید از ابعاد مختلف مورد بررسی قرار گیرد، بنابراین علم‌سنجی سعی دارد با استفاده از داده‌های کمّی مربوط به تولید، توزیع و استفاده از متون علمی، علم و پژوهش‌های علمی را توصیف و ویژگی‌های آن را مشخص کند (حیدری، ۱۳۸۹).

«ترسیم نقشه علم»[3]، «ترسیم نقشه دانش»[4]، «مصورسازی علم»[5] و «مصورسازی دانش»[6]، نام‌های گوناگونی از حوزه میان‌رشته‌ای جدید هستند که در سال‌های اخیر در علم‌سنجی، کاربردی وسیع یافته‌اند؛ این حوزه، از طریق پردازش، استخراج و مرتب‌سازی اطلاعات به ترسیم نقشه علم می‌پردازد و امکان تحلیل، مسیریابی و نمایش دانش را فراهم می‌آورد؛ علاوه‌بر آن، به منظور سهولت بخشیدن دسترسی به اطلاعات، آشکارسازی ساختار دانش و کمک به جستجوگران دانش برای رسیدن به نتایج موفقیت‌آمیز حرکت می‌کند (Borner, 2004 Shiffrin &). البته لازم به یادآوری

1. Positivism
2. Scientometrics
3. Mapping Science
4. Mapping Knowledge
5. Visualization of Science/ Visualizing the Science
6. Visualization of Knowledge/ Visualizing the Knowledge

است که مصورسازی به صورت کلی در دو محدوده علمی قابل تعریف است: «مصورسازی علمی»[1] و «مصورسازی اطلاعات»[2]. این تمایز به صورت تجربی در ساختار متفاوت آنها نسبت به ورود داده‌ها و یا داده‌های خامی قابل مشاهده است که به شکل تصویری درآمده‌اند (Polanco & Zartal, 1999). در واقع، مصورسازی اطلاعات یکی از حوزه‌های علوم رایانه‌ای است که با نوآوری در ارائه مقادیر عظیم اطلاعات ارتباط دارد (Walter, Stuart & Borisyuk, 2004). به این معنا که با تکیه بر چنین شیوه‌هایی می‌توان میزان شناخت مخاطب را افزایش داد و شرایط بهتری برای تعامل با عناصر اطلاعاتی برقرار کرد. رابل و همکاران[3] (۲۰۱۰) نیز اظهار می‌دارند که مصورسازی علمی از تحلیل جزئی داده‌های فیزیکی پشتیبانی می‌کند، ولی مصورسازی اطلاعات به اکتشاف فضاهای متغیر و تعیین روابط بین ابعاد مختلف داده‌ها می‌پردازد. آگاتر و برمودز[4] (۲۰۰۵) نیز به زمینه‌های همکاری بین‌رشته‌ای مصورسازی توجه می‌کنند و مطالعات اصلی آن را در میان حوزه‌های مطالعاتی مربوط به هنر، طراحی، علم و فناوری برمی‌شمارند.

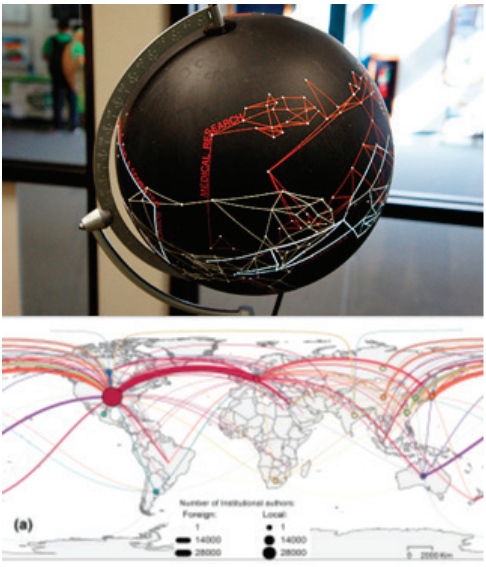
اهمیت مصورسازی علم و تأثیر شگرفی که می‌تواند بر درک بهتر روابط موجود در علم داشته باشد، عده‌ای را به این باور رسانیده که با کمک این حوزه، می‌توان «دانش ضمنی»[5] را، که در گذشته بیشتر برای دانشمندان هر حوزه درک شدنی بود، به صورت آشکار نمایش داد و آن را به «دانش آشکار»[6] تبدیل کرد (نوروزی چاکلی، ۱۳۹۰). بر مبنای این ضرورت دانشمندان حوزه مطالعات علم به دنبال رویکردی نوین در راستای شناسایی این ساختارها بوده و هستند. رویکردی که با عنوان مصورسازی حوزه دانش، مسیر بلوغ خود را می‌پیماید. این رویکرد با اتکای صرف به «تحلیل هم‌استنادی»[7] آغاز و رفته‌رفته با برقراری پیوند با فنون «متن‌کاوی»[8] و مصورسازی در راستای تکامل، پیش رفته است (زوارقی، فدایی و فهیم‌نیا، ۱۳۹۰). بر همین اساس، در این مقاله سعی شده است، نگاهی کلی به این حوزه پژوهشی نوپدید مطرح و در زمینه اصلی‌ترین مفاهیم موجود در این زمینه مباحثی ارائه شود.

چیستی نقشه‌های علمی

همانگونه که نقشه‌های جغرافیایی قرن‌هاست در اکتشاف و مسیریابی[9] به ما کمک کرده‌اند، نقشه‌های

1. Scientific Visualization
2. Information Visualization (IV)
3. Rübel and et al.
4. Agutter & Bermudez
5. Implicit Knowledge
6. Explicit Knowledge
7. Co-citations Analysis
8. Text-Mining
9. Navigate

علمی نیز با سابقه اندک، از مطالعات اُتله[1] (۱۹۱۸، به‌نقل از نوروزی چاکلی، ۱۳۹۰) آغاز و تاکنون با دستاوردهای دانشمندان این حوزه (هرچند با بعضی نقصان‌ها در زمینه تفسیر و تبیین) همراه شده است (زوارقی، فدایی و فهیم‌نیا، ۱۳۹۰). مراد از «نقشه» در اصطلاح «نقشه علم»، نگاشت به معنایی نیست که در ترکیب «ره‌نگاشت»[2] به‌کار می‌رود، بلکه نقشه علم همانند یک نقشه جغرافیایی است. همانگونه که در یک نقشه جغرافیایی می‌توان موقعیت شهرها را دید و بزرگی و کوچکی آنها نسبت به هم را مشاهده کرد و دوری و نزدیکی آنها از یکدیگر، همسایگی و راه‌های ارتباطی آنها را تشخیص داد، در نقشه‌های علمی نیز می‌توان در مورد وضعیت مفاهیم – مانند شهرها و به همان کیفیت– اطلاع کسب کرد (ناصری جزه، طباطبائیان و فاتح راد، ۱۳۹۱). یک نقشه علمی نمایش‌دهنده فضای چگونگی ارتباط رشته‌ها، حوزه‌ها، تخصص‌ها و مقاله‌های فردی یا گروهی نویسندگان با یکدیگر است که از طریق نزدیکی فیزیکی یا موقعیت‌های نسبی نشان داده می‌شوند؛ همانند نقشه‌های جغرافیایی که نشان‌دهنده روابط سیاسی یا ویژگی‌های فیزیکی زمین هستند (Small, 1999).



شکل ۱. نمونه‌هایی از نقشه‌های علمی به شکل کره جغرافیایی زمین و نقشه‌های مسطح جغرافیایی

مویا[3] معتقد است، که ترسیم اطلاعات علمی در قالب نقشه‌های علمی یک نوع تفسیر است، به طوری که داده‌ها و پدیده‌های پیچیده واقعی به پیام‌های قابل درک تبدیل می‌شوند و این ترسیم باعث می‌شود که داده‌ها و پدیده‌های پیکره نامرئی ساختار دانش برای ما قابل درک شود (Small, 2000).

1. Otlet
2. Road Mapping
3. Moya

در واقع مصورسازی اطلاعات یکی از روش‌هایی است که به منظور انتقال بهتر اطلاعات و بهره‌گیری مناسب از روش‌های نمایش آن، مورد استفاده قرار می‌گیرد. با توجه به اینکه در این شیوه، تبادل و انتقال اطلاعات به صورت دیداری انجام می‌شود، تلاش می‌شود تا با نمایش تصویری اطلاعات قدرت درک و یادگیری کاربر افزایش یابد و حجم زیادی از اطلاعات به صورت فشرده، با حجمی کمتر و به صورت مصور ارائه شود. به عبارت دیگر، شعار نهفته در مصورسازی اطلاعات علمی، به کارگیری نحوه نگاه یا بینش برای تفکر است. واضح است که مصورسازی اطلاعات علمی زمانی اهمیت دارد که با حجم زیادی از اطلاعات روبه‌رو باشیم (درودی، ۱۳۸۷).

به عقیده نایونز[1] (۱۹۹۹)، نقشه علمی[2] عبارت است از تجزیه و تحلیل انتشارات یک حوزه علمی از زوایای مختلف و ترسیم نگرشی کلی از آن حوزه که بر پایه این نقشه و ترسیم سیر تغییر و تحولات، حوزه‌هایی که بیشترین و کمترین نزدیکی را دارند از هم متمایز می‌شوند. به این ترتیب، هر کاربر افزون بر ویژگی‌ها و ارتباطات بین زیرحوزه‌های هر حوزه از علم می‌تواند تأثیرگذارترین افراد و مؤسسه‌های تحقیقاتی را نیز در آن حوزه خاص مشخص کند؛ و در ادامه بیان می‌دارد که هدف از تهیه نقشه علمی، شناسایی نقاطی از دانش است که به اصطلاح، «بحث داغ»[3] حوزه مربوط به خود را پیگیری می‌کنند. در نقشه‌های علم، حوزه‌های موضوعی که با هم ارتباط بیشتری دارند در فاصله نزدیک‌تر و حوزه‌هایی که ارتباط کمتری دارند در فاصله بیشتری نمایش داده می‌شوند.

این نوع نقشه‌ها تلاش دارند تا فرایندهای رشد، ادغام و تجزیه حوزه‌های مختلف علم طی زمان را نشان دهند. این رویکرد، علم را به عنوان تحقیقی در صفحه مسائل تجریدی، مفهوم‌سازی می‌کند. حوزه‌های علمی در این نقشه‌ها، به نسبت میزان فعالیت دانشمندان در آنها مشخص می‌شوند و فضاهای خالی نشان‌دهنده حوزه‌های کار نشده و یا ناشناخته علم است. در این نوع نمایش، می‌توان رشد، ادغام و یا تجزیه حوزه‌های مختلف علمی را طی زمان رصد کرد (Börner & Scharnhorst, 2009).

کاربرد نقشه‌های مفهومی را می‌توان در دو مورد خلاصه کرد: نشان دادن دینامیک‌های کمی گروهی از مفاهیم در یک حوزه علمی (که در نقشه، تشکیل یک خوشه را می‌دهند) و همچنین کشف روابط بین مفاهیم (Marshakova-Shaikevich, 2005)

به عقیده بورنر (نقل در محمدی، ۱۳۸۷) پژوهشگران یک حوزه می‌توانند با استفاده از ترسیم یک حوزه از علم به راحتی به نتایج پژوهش‌ها دست یابند. افراد کلیدی هر حوزه را شناسایی کنند و با آنها در زمینه‌های مشترک همکاری علمی داشته باشند. مهمتر از همه، به کمک نقشه‌های یک حوزه از علم راحت‌تر می‌توان از تجربه پژوهش‌های پیشین استفاده کرد و در زمان پژوهش صرفه‌جویی

1. Noyons

2. Map of Science

۳. در حوزه‌های داغ، فعالیت علمی بیشتری صورت می‌گیرد و در مصورسازی علوم، بیشتر به صورت گراف‌های مرکزی و بزرگ به نمایش درمی‌آیند.

کرد؛ برای ناشران، چون در نقشه‌های علمی همه اطلاعات یک حوزه خاص، یکجا ارائه می‌شود، این یک‌دستی در ارائه اطلاعات باعث می‌شود که کاربران بیشتری به آسانی از اطلاعات ناشران استفاده کنند؛ برای جوامع علمی، یک جامعه علمی که در حوزه‌های تخصصی علم فعالیت می‌کند، به کمک نقشه‌های علمیِ آن حوزه‌ها می‌تواند علاوه بر اینکه به‌راحتی به اطلاعات گذشته دسترسی پیدا کند، برای آینده حوزه‌های مورد پژوهش خود برنامه‌ریزی کند، که این امر باعث تقویت دانش در آن جامعه علمی می‌شود؛ برای مؤسسه‌های سرمایه‌گذار، نقشه‌های علمی برای کنترل درازمدت چرخش سرمایه و توسعه پژوهش، ارزیابی سیاست‌گذاری برای برنامه‌های مختلف، تصمیم‌گیری برای زمان انجام یک پروژه و الگوی سرمایه‌گذاری علمی می‌تواند برای مؤسسه‌های فعال در این زمینه سودمند باشد. همچنین می‌توان از نیروهای انسانی در راستای توسعه برنامه علمی استفاده کرد. به‌علاوه، می‌توان در زمینه مشخص کردن حوزه‌های پژوهشی آتی و شبیه‌سازی حوزه‌های پژوهشی جدید، از نقشه‌های علمی استفاده کرد؛ برای صنعت، با توجه به اهمیت ارتباط صنعت و دانشگاه، مدیران بخش صنعت با استفاده از نقشه‌های علمی قادرند از نتایج اصلی پژوهش‌های مرتبط، به راحتی و در زمان کم آگاه شوند. همچنین می‌توانند با کمک نقشه‌های علمی نیازهای خود را شناسایی کنند و در قالب طرح‌های پژوهشی به پژوهشگران سفارش دهند یا اینکه پژوهش‌ها را در مسیر کاربردی بودن، هدایت کنند. در شکل (۲) نمونه‌ای از کاربردهای نقشه‌های علمی آورده شده است.

شکل ۲. استفاده از نقشه‌های علمی به‌عنوان قطار ۳۰۰ متری علم در کشور آلمان برای بازدید سیاستمداران کشوری و لشکری، استفاده در کلاس‌های درس، موزه‌های علم و بازدید کودکان و دانش‌آموزان و استفاده توسط دانشمندان حوزه‌های مختلف علوم

روش‌های ترسیم نقشه علم

در ترسیم ساختار علم سه جزء در نظر گرفته می‌شوند: عناصر فردی، عناصر مرتبط با یکدیگر که

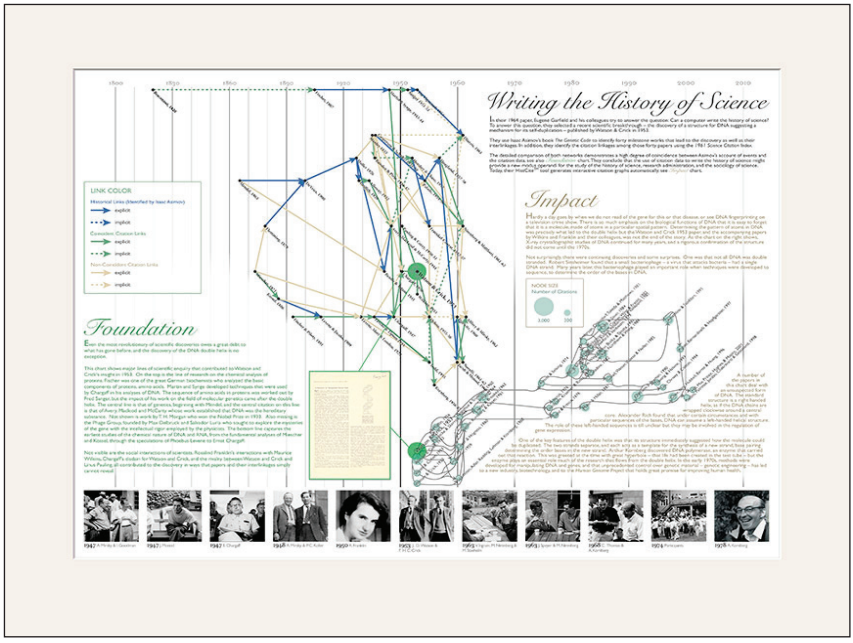
یک شبکه را به وجود آورده‌اند و تفسیر روابط بین عناصر (پشوتنی‌زاده و عصاره، ۱۳۸۸). همانطور که پیش‌تر سخن رفت، به طور معمول کشورها، مؤسسه‌ها، مؤلفان، مدارک، حوزه‌های علمی و واژه‌ها، واحدهای تحلیل در نقشه‌های علمی هستند. برای ترسیم نقشه‌ها باید به نوع روابط میان موجودیت‌ها دقت کرد، بنابراین باید مقیاسی وجود داشته باشد تا میزان روابط میان موجودیت‌ها را بسنجد. به عبارتی، باید دید دو موجودیت تا چه اندازه با هم مرتبط هستند، درک این ارتباط نیازمند یک مقیاس است. ذکر این نکته نیز ضروری است که مقیاس اندازه‌گیریِ ارتباط برای واحدهای تحلیل (موجودیت‌ها) ممکن است متفاوت باشد (Van Eck & Waltman, 2007).

رابطه مجله‌ها، مدارک و مؤلفان اغلب بر اساس ارتباط استنادی بررسی می‌شوند. «استناد یا پیوند مستقیم»[1]، «هم‌استنادی»[2]، و «زوج‌های کتابشناختی»[3] سه مقیاس مشهور برای سنجش ارتباط استنادی میان موجودیت‌ها هستند (Small, 1999). روش دیگر برای سنجش ارتباط میان داده‌های کتاب‌شناختی، روابط «هم‌تألیفی»[4] است(White & Griffith, 1981; White & McCain, 1998). این مقیاس می‌تواند ارتباط میان مؤلفان، مؤسسه‌های علمی و کشورها را تعیین کند؛ و این ارتباط بر اساس تعداد تألیفات مشترک میان آنها سنجیده می‌شود. روش‌های معمول دیگر ترسیم نقشه‌ها، «هم‌رخدادی»[5]، «هم‌رخدادی واژگان»[6]، «همبستگی میان حوزه‌های موضوعی»[7] یا مواردی دیگر از این قبیل است که برای سنجش ارتباط میان مدارک و مؤلفان آنهاست.(Callon, Courtial & Laville, 1991;Shiffrin & Borner, 2004)

در دایرةالمعارف کتابداری و اطلاع‌رسانی «استناد» چنین تعریف شده است: «استناد به معنای قرار دادن چیزی، تکیه بر چیزی کردن و آیه، حدیث، یا سخنی را سند قرار دادن و بدان تمسّک جستن آمده و در اصطلاح، اشاره به سخن یا سند پیشین را گویند». معمولاً پژوهشگران در نوشته‌های خود به آثاری ارجاع می‌دهند که ربط موضوعی با نوشته آنها دارد و از این آثار برای تأیید نظر خویش استفاده می‌کنند یا تفاوت نظر خود را با اندیشه‌ها و یافته‌های جدید را نشان می‌دهند. این آثار را، که به منابع یا مأخذ نیز شهرت دارد، استناد یا سند[8] و نوشته‌ای که به آنها استناد می‌کند، استنادکننده متن[9] می‌نامند (مدیر امانی،۱۳۸۱). تحلیل استنادی، نه تنها اقدام به بررسی چگونگی نشر و استفاده از انواع منابع

1. Direct Citaition-link
2. Co-Citaition
3. Bibliographic Coupling
4. Co-Authorship
5. Co-Occurring
6. Co-Words
7. Co-Classification
8. Cited Source
9. Citing Source

اطلاعاتی می‌کند، بلکه به کشف رابطه بین متن و مأخذ و نیز تعیین ویژگی‌های یک زمینه موضوعی از طریق بررسی دقیق در متون آن حوزه می‌پردازد (Lowe, 2003). به سخن دیگر، در این تحقیقات فقط به رابطه میان سند و متن توجه نشده بلکه بر مبنای این رابطه و با توجه به فرض اینگونه تحلیل‌ها، روابط میان خود «متن»ها و یا خود «سند»ها نیز موضوع تحقیقات بسیاری است و اساس راه‌حل‌هایی برای عمل گزینش متون بنیانی و کشف معتبر از نامعتبر قرار گرفته است. به همین سبب و برای سهولت بررسی اینگونه پژوهش‌ها، شاید بتوان آنها را از لحاظ شیوه عمل به دو گروه عمده تحلیل استنادی عمودی و تحلیل استنادی افقی تقسیم کرد (حرّی، ۱۳۸۱).

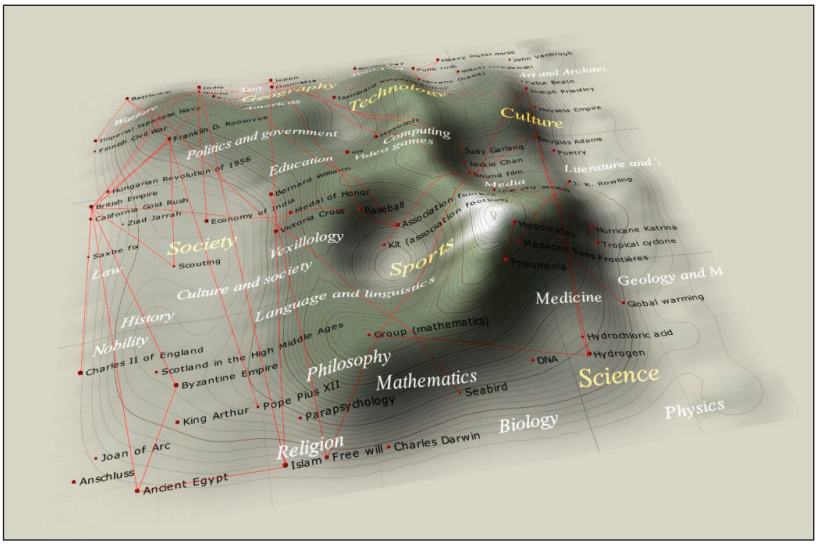


شکل ۳. نمونه‌ای از نقشه‌های استنادی علوم برای دانشمندان

در پاسخ به پیشرفت‌های صورت گرفته در زمینه نقشه‌های هم‌استنادی در سال‌های دهه ۱۹۷۰ توسط اسمال و گریفیث[1] (۱۹۷۳؛ ۱۹۷۴) افرادی از قبیل کالون، کورتیال، تورنر و باواین[2] در فرانسه طرح نقشه‌های هم‌رخدادی واژگان را به عنوان زمینه تحقیقاتی دیگری برای مطالعه روابط معنایی در ادبیات تحقیق علم و فناوری مطرح ساختند و آن را «لگزی مَپ»[3] نام نهادند (Callon et al., 1983; Callon, Law & Rip, 1986)؛ از آن زمان تاکنون این فنون برای «ترسیم

1. Small & Griffith
2. Callon, Courtial, Turner & Bauin
3. Leximappe

نقشه هم‌رخدادی واژگان» توسعه بیشتری یافتند. هم‌رخدادی واژگان به این معنی است که دو واژه با همدیگر در بعضی بخش‌های زبان طبیعی مثل عنوان یا چکیده ظاهر شوند. این میزان هم‌رخدادی توسط کامپیوتر قابل محاسبه است و از این میزان هم‌رخدادی داده‌های خام برای ترسیم یک حوزه علمی استفاده می‌شود(Börner, Chen & Boyack, 2003). در گذشته هم‌رخدادی واژگان به عنوان روشی برای توسعه اهداف سیاسی مورد استفاده قرار می‌گرفت (Noyons, 2001). بر اساس تجزیه و تحلیل هم‌رخدادی واژگان[1] می‌توان موضوع‌های علم را استخراج و ارتباط بین این موضوع‌ها را به صورت مستقیم از محتوای موضوعی متون علمی و پروانه‌های ثبت اختراع[2] کشف کرد (Callon, Law & Rip, 1986; Li et al., 2009). برجستگی‌های شکل (۴) نشاندهنده میزان چگالی بیشتر موضوع‌ها و خطوط قرمز ارتباط بین مقاله‌ها را نشان می‌دهد.

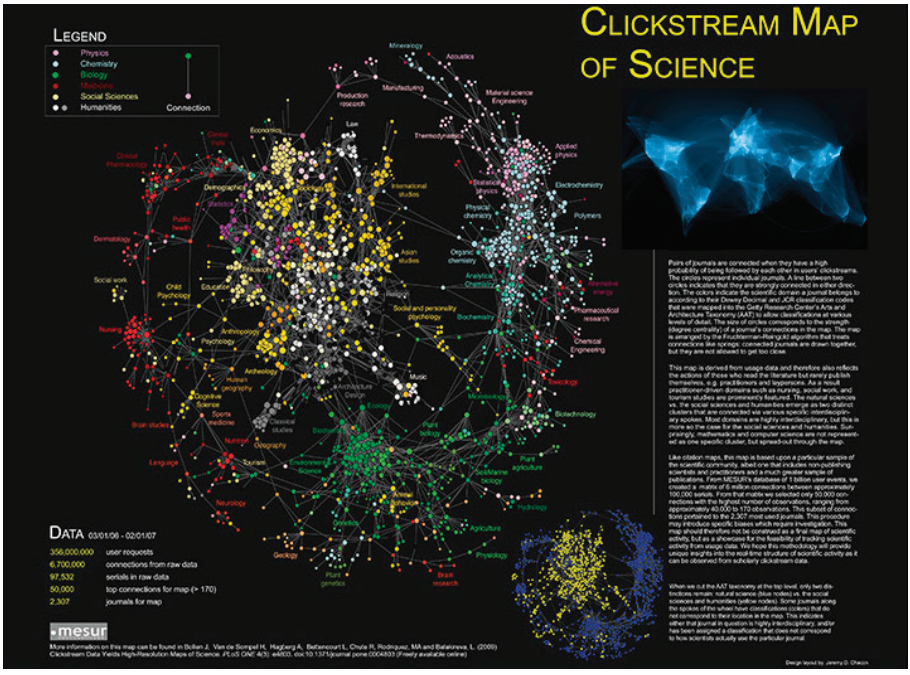


شکل ۴. نمایش رخداد موضوعی مقاله‌های موجود در ویکی‌پدیا (بر اساس بسامد واژگان)

در محیط‌های پژوهشی و بحث دربـاره همکاری‌های علمی، هم‌تألیفی رؤیت‌پذیرترین، دسترس‌پذیرترین و ارزان‌ترین شاخصی است که در راستای سنجش و اندازه‌گیری میزان همکاری‌های علمی به‌کار می‌رود. محاسبه هم‌تألیفی در انتشارات علمی به لحاظ نظری ساده و به طور محسوسی با میزان همکاری‌های علمی در ارتباط است (Lundberg et al., 2006; Harsanyi, 1993; Katz & Martin, 1997; Cronin, Shaw & La Barre, 2004; Wang et al., 2005). به باور چئونگ و

1. Co-word Analysis
2. Patents

کوربیت[1] (۲۰۰۹) تاکنون در مطالعات متعددی بر وجود همبستگی مثبت میان همکاری‌های علمی و هم‌تألیفی تأکید داشته‌اند که از این منظر می‌توان هم‌تألیفی را یکی از ملموس‌ترین و قابل استنادترین اشکال همکاری‌های پژوهشی در نظر گرفت. پدیده «تألیف مشترک» یا «همکاری در تألیف» یا «هم‌نویسندگی/هم‌تألیفی» که برابر نهاده‌های واژه «Co-authorship» یا «Joint authorship» در زبان انگلیسی است (رحیمی و فتاحی، ۱۳۸۶)؛ تنها یکی از شاخص‌های سنجش همکاری علمی است که در تولیدات علمی نظیر کتاب، مقاله، پژوهش، و امثال آنها منعکس می‌شود (Harsanyi, 1993؛ عطاپور، حسن‌زاده و نوروزی چاکلی، ۱۳۸۸)؛ و به عنوان رسمی‌ترین جلوه همکاری فکری میان نویسندگان در تولید پژوهش‌های علمی، عبارت است از مشارکت دو یا چند نویسنده در تولید یک اثر که به تولید بروندادی علمی با کمیت و کیفیت بالاتری در مقایسه با زمانی منجر می‌شود که یک فرد به تنهایی اثری را تولید و منتشر کند (Hudson, 1996).



شکل ۵. نمونه‌ای از گراف‌های همکاری علمی در بین رشته‌های علمی در سطح جهانی

با توجه به مطالبی که ارائه شد، نقشه دانش یک رشته را می‌توان با روش‌های «تحلیل هم‌استنادی» و یا روش‌های «تحلیل هم‌رخدادی واژگان» و البته با «تحلیل همکاری علمی»[2] ترسیم

1. Cheong & Corbit
2. Scientific Collaboration Analysis

کرد. به هر حال، امروزه پرکاربردترینِ روش‌ها برای ترسیم نقشه‌های مفهومی، تحلیل هم‌رخدادی واژگان است. علت این که تمایلات به تحلیل‌های هم‌رخدادی نسبت به تحلیل‌های هم‌استنادی بیشتر شده است را می‌توان دو چیز دانست؛ از نظر علمی، در روش هم‌رخدادی واژگان، می‌توان مدارکی که به آنها استناد نشده است را نیز در تحلیل وارد کرد[1]. همچنین از نظر روش، در تحلیل‌های هم‌استنادی، تحلیل نحوه پویایی (دینامیک) حوزه مورد پژوهش با روندهای موجود در عمل نویسندگان مقاله‌ها، آمیخته می‌شود و برآورد خالصی از پیشرفت علم به ما نمی‌دهد (Yoon & Park, 2004; Janssens et al., 2006). تفاوت مهم دیگری که هه[2] (۱۹۹۹) به آن اشاره کرده است اینکه، در یک دوره پژوهشی مشخص، تحلیل هم‌استنادی به منابع استناددهنده (مقاله استناددهنده، مؤلفان استناددهنده) و مأخذ استنادشده (مؤلف استنادشده، مدرک استنادشده) نیازمند است؛ اما تحلیل هم‌واژگانی فقط نیازمند مجموعه‌ای از مقاله‌های مجله‌ها (مدارک) در یک حوزه موضوعی خاص است.

با توجه به مطالب پیش‌گفته، می‌توان بیان داشت که نقشه علم یکی از خروجی‌هایی است که «مطالعات کمی علم و فناوری»[3] برای کمک به سیاست‌گذاران در دسترس قرار می‌دهند؛ با این رویکرد می‌توان بیان داشت که نقشه، ارتباط ایستای اجزای یک نظام را نشان می‌دهد و «نقشه دانشی»، قادر است تا منابع و مسیر جریان دانش، و محدودیت‌ها و کمبودهای دانش را مشخص کند و با احصاء حوزه‌های اصلی آن دانش، اطلاعات لازم در مورد هر زیرحوزه را در اختیار مدیران قرار دهد. زمانی که نقشه دانشی در مورد یک رشته علمی ترسیم می‌شود، یک «نقشه علم» حاصل می‌شود. «نقشه علم»، تعداد زیرحوزه‌های هر زمینه علمی و میزان دانش موجود در هر زیرحوزه، و نیز ارتباط و تعامل زیرحوزه‌های مختلف با یکدیگر را مشخص می‌کند.

نکته قابل ذکر اینکه در ترسیم نقشه‌های علمی؛ نقشه به عنوان یک عبارت تنها به بخش‌های ترسیم شده برنمی‌گردد، بلکه گویای فنون تجزیه و تحلیل نیز هست. همانگونه که بویاک اشاره می‌کند، تجزیه و تحلیل نقشه دانش یک حوزه، علاوه بر تکنیک‌های ترسیم اطلاعات می‌تواند شامل موضوع‌های گوناگونی مثل تجزیه و تحلیل شبکه (وب، شبکه‌های اجتماعی[4] و شبکه‌های بزرگ)، زبان‌شناسی، استخراج مفاهیم و موضوع‌ها، تحلیل استنادی و شاخص‌های علم و فناوری آن حوزه باشد (Boyack, 2004). همچنین، با وجود اهمیت مصورسازی در علم‌سنجی، برای اینکه نقشه‌های علم بتوانند نتایجی معنادار و حائز اهمیت را ارائه کنند، باید از روش‌های صحیح بهره گیرند و داده‌های معتبر و قابل اعتماد را نیز مبنا قرار دهند؛ از این‌رو، با وجود ظهور ابزارهای گوناگون برای ترسیم نقشه‌های علمی، تمامی این ابزارها هنگامی می‌توانند نقش مؤثری را در ارائه اطلاعات صحیح ایفا کنند و زمانی می‌توانند وظیفه خود را به درستی به انجام برسانند که داده‌های معتبر و باکفایتی را در

۱. شامل منابعی که هرگز استناد دریافت نخواهند کرد و منابع بالقوه‌ای که در بازه زمانی پژوهش استناد دریافت نکرده‌اند.

2. He

3. S & T Metrics

4. Social Network

اختیار داشته باشند (نوروزی چاکلی، ۱۳۹۰).

فرایند مفهومی ترسیم نقشه‌های علم

فرایند مفهومی ترسیم نقشه‌های موضوعی علوم مبتنی بر نظر بورنر و همکارانش، شامل شش مرحله است که عبارتند از:

نخستین گام در هر فرایند نگاشت یا ترسیم نقشه؛ استخراج اطلاعات مناسب است. در این مرحله استراتژی‌های مختلف جستجو کاربرد دارند؛

انتخاب واحدهای تحلیل؛ بستگی به سؤالی دارد که درصدد پاسخگوئی به آن هستیم. رایج‌ترین واحدها برای ترسیم نقشه‌ها، نوشته‌ها هستند که عبارتند از: مجله‌ها، مدارک، نویسندگان، واژگان و اصطلاحات توصیفگر؛

واژه‌های تکنیکی بسیاری به عنوان شاخص‌های شناسایی شباهت بین مقاله‌های به‌کار برده می‌شوند، این واژه‌ها از پیشوندهای Inter و Co ساخته شده‌اند؛

شباهت‌های بین مدارک (واحدها) معمولاً با روش‌های مختلفی محاسبه می‌شوند که رایج‌ترین آنها عبارتند از: ارتباطات استنادی یا ارجاعی[۱]، شباهت‌های هم‌رخدادی[۲]، مدل بردار فضایی[۳]؛

روش‌های دسته‌بندی متنوعی با توجه به کاربرد هر یک در ترسیم نقشه‌ها وجود دارند که مهمترین آنها تجزیه مقدار ویژه/بردار ویژه[۴]، تحلیل عاملی[۵]، مقیاس‌بندی چندبعدی[۶]، تحلیل معنایی نهفته[۷]، تحلیل خوشه‌ای[۸] و مثلث‌بندی[۹] هستند؛

در آخرین مرحله نوبت به استفاده از فنون نمایش اطلاعات در قالب بصری می‌رسد. نمایش[۱۰] به تمام روش‌های مصورسازی اطلاعات گفته می‌شود که در راستای جستجو و پیمایش اثربخشی فضاهای گسترده اطلاعاتی هستند. از جمله این روش‌ها می‌توان به انواع روش‌های پالایش کردن[۱۱] اطلاعات، انواع روش‌های بزرگ‌نمایی[۱۲] و تغییر زاویه[۱۳] دید اشاره کرد(Börner, Chen & Boyack, 2003) .

1. Citation Linkages
2. Co-occurrence Similarities
3. Vector Space Model
4. Eigen value/Eigenvector Decomposition
5. Factor Analysis
6. Multidimensional Scaling
7. Latent Semantic Analysis
8. Cluster Analysis
9. Triangulation
10. Display
11. Filtering
12. Zooming
13. Distortion

جدول ۱. فرایند ترسیم نقشه‌های علمی(Börner, Chen & Boyack, 2003)

استخراج داده[۴]	واحدهای تحلیل	سنجه‌ها[۳]	ترسیم[۲] (کاهش ابعاد و طرح‌بندی)		نمایش[۱]
			شباهت بین واحدها[۶]	طرح‌بندی[۵]	
منابع جستجو * آی اِس آی * اینسپک[۸۲] * نمایه‌نامه مهندسی[۹۲] * مدلاین[۰۳] * نمایه پژوهش[۱۳] * ثبت اختراع‌ها[۳۳] و غیره **شیوه مرزبندی** * توسط استنادها[۳۳] * توسط اصطلاحات[۳۴]	**انتخاب‌های معمول** * کشورها * حوزه‌های موضوعی * مجلات * مدارک * نویسندگان * اصطلاحات	**شمار/ فراوانی** * خصیصه‌ها (مثل اصطلاحات) * استنادات مؤلفان * هم‌استنادی‌ها * برحسب سال[۵۲] **آستانه[۶۲]** * برحسب شمارها[۲۷]	**کمیت (واحد براساس موجودیت‌های ماتریس)[۴۱]** * استناد مستقیم[۵۱] * هم‌استنادها * پیوندهای ترکیب شده (زوج‌های کتاب‌شناختی)[۶۱] * هم‌رخدادی واژگان[۷۱]/ هم‌رخدادی اصطلاحات[۸۱] * هم‌رده‌ای[۹۱] **بردار (واحد بر اساس موجودیت‌های ماتریس)[۰۲]** * مدل بردار فضایی[۱۲] (واژگان/ اصطلاحات) * تحلیل معنایی ضمنی (واژگان/ اصطلاحات) شامل فنون تجزیه- مقدار (اِس وی دی)[۲۲] **همبستگی (در صورت نیاز)[۳۲]** * همبستگی پیرسون[۳۴] در مورد هر یک از موارد فوق	**کاهش ابعاد** * بردار ویژه/ روش‌های ویژه‌مقدار * تحلیل عاملی و تحلیل مؤلفه‌های اصلی[۲۱] * مقیاس‌بندی چندبُعدی * شبکه‌های مسیریاب * نقشه‌های خودسازمانده شامل اِس اُ. اِم، ائی تی- مَپ و غیره **تحلیل خوشه‌ای روش‌های کمی** * مثلث‌بندی * جای‌گذاری نیرو- هدایت شونده[۱۳]	**تعامل[۷]** * گشتن[۸] * استخراج[۹] * بزرگ‌نمایی * پالایش * پرس‌وجو[۰۱] * جزئیات[۱۱] **تحلیل**

ساختار درونی علم و فرایند کشف دانش

مطابق با مطالعات ناگ پاول سه موضوع در مطالعات علم‌سنجی پیگیری شده است؛ علم‌سنجی و سیاست‌های علمی و فناورانه (شامل روش‌ها و حوزه‌های علم‌سنجی)، ساختار و پویایی‌های علم (شامل کارهای فردی تا همکاری‌های بین‌المللی دانشمندان) و جنبه‌های منطقه‌ای علوم. بر اساس طبقه‌بندی فوق یکی از موضوع‌های علم‌سنجی مطالعه ساختار علم و پویایی آن است. به این مفهوم که در درجه اول برای کل دانش و درجه بعدی برای هر یک از رشته‌های مختلف ساختار و حوزه‌های تخصصی مشخص می‌شوند (Hood & Wilson, 2001). در واقع توسعه محتوا و ساختار علم، اساسی‌ترین ویژگی مربوط به بعد شناختی هر علمی است (Moed, Glanzel & Schmoch, 2004). برای درک این شناخت باید دانش مفید حوزه‌های علمی را استخراج کرد. در تعریفی که توربن

و همکارانش (۱۹۹۹) از کشف یا استخراج دانش دارند این است که آن را فرایند استخراج دانش مفید از میان انبوهی از داده‌ها دانسته‌اند. کشف یا استخراج دانش به دو صورت است:

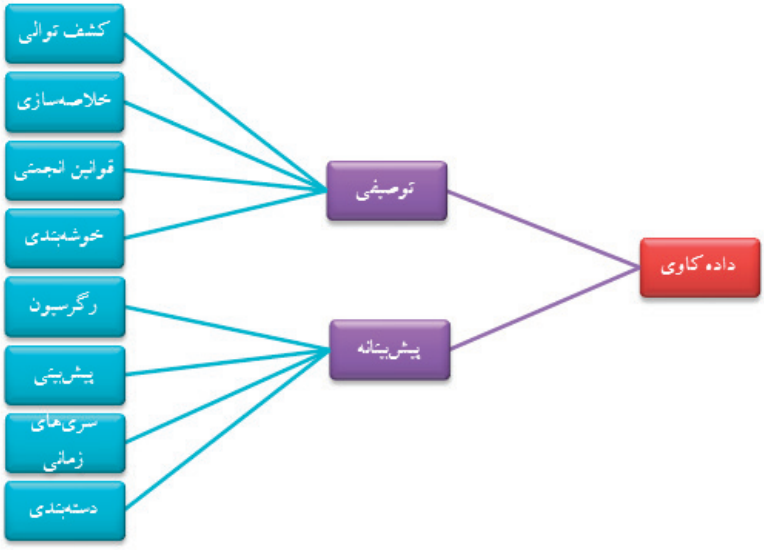
الف) استخراج دانش از پایگاه داده یا داده‌کاوی؛

ب) استخراج دانش از متن‌ها یا متن‌کاوی (Hotho, Nürnberger & Paaß, 2005).

لازم به ذکر است که پایگاه داده‌ها به صورت ساخت‌یافته ، برای تحلیل مقادیر زیادی از داده‌ها و برنامه‌های پردازش خودکار، طراحی شده‌اند؛ در حالی که متن، بی‌شکل، بدون ساختار و برای خواندن مردم نوشته شده است و با وجود تنظیم مشکل آن، رایج‌ترین وسیله برای تبادل رسمی اطلاعات است (Taylor, 2006).

داده‌کاوی

داده‌کاوی، حوزه‌ای میان‌رشته‌ای است که حوزه‌های مختلفی همچون پایگاه داده، آمار، یادگیری ماشینی و سایر زمینه‌های مرتبط را با هم تلفیق می‌کند (Han & Kamber, 2006). روش‌های اصلی داده‌کاوی دو دسته هستند: توصیفی و پیش‌بینانه . وظایف روش‌های توصیفی، تشخیص خواص عمومی داده و هدف روش‌های توصیفی، یافتن الگوهایی در مورد داده‌هاست که برای انسان قابل تفسیر باشد. وظایف پیش‌بینانه، به منظور پیش‌بینی رفتارهای آینده آنها استفاده می‌شود. منظور از پیش‌بینی به‌کارگیری چند متغیر یا فیلد در پایگاه داده برای پیش‌بینی مقادیر آینده یا ناشناخته دیگر متغیرهای مورد علاقه است (غضنفری، علیزاده و تیمورپور، ۱۳۸۷). در شکل (۶) عملکردهای داده‌کاوی نشان داده شده است.



شکل ٦. عملکردهای داده‌کاوی (Duan, 2011)

به‌عبارت دیگر، داده‌کاوی یا استخراج دانش از پایگاه داده‌ها[1]، فرایند مهم شناسایی الگوهای معتبر، جدید و قابل فهم در میان انبوهی از داده‌ها است. مفهوم داده‌کاوی شامل الگوریتم‌ها و روش‌هایی است که باعث استخراج اطلاعات از داده‌ها می‌شود (Fayyad, Piatetsky-Shapiro & Smyth, 1996) که در اینجا به شرح آنها می‌پردازیم. مطابق با شکل (۷)، فرایند داده‌کاوی به شرح زیر است:

آماده کردن داده‌ها؛ که به صورت انتقال داده‌های جمع‌آوری شده از منابع مختلف داخلی و خارجی به انباره داده‌ها است؛

انتخاب مجموعه داده‌های لازم و معنادار برای کاوش و همچنین پاک‌سازی و پردازش آنها؛ یعنی اصلاح خطاها یا تناقض‌های بین آنها مثل اصلاح خطاهای نوشتاری، اجتناب از تکرار غیر ضروری داده‌ها و کنترل برای همسان کردن داده‌ها از نظر شکل؛

تغییر شکل داده‌ها؛ یعنی گروه‌بندی یا خلاصه کردن داده‌ها؛

انتخاب روش‌های خاص داده‌کاوی؛ با توجه به اهداف داده‌کاوی (یعنی پیش‌بینی و توصیف)، که این روش‌ها عبارتند از:

طبقه‌بندی (دسته‌بندی)؛ در این شیوه، اطلاعات مورد نظر در گروه‌های از پیش تعریف شده قرار گرفته و رده آنها مشخص می‌شود؛

گروه‌بندی (خوشه‌بندی)؛ در این شیوه، اطلاعات و مجموعه داده‌های وارد شده به گروه‌های مشابه بر اساس فراوانی نسبی، نه بر اساس ویژگی‌هایِ‌شان طبقه‌بندی می‌شوند؛

رگرسیون؛ در این روش بر اساس داده‌های ورودی و خطوط تصمیم‌گیری، داده‌های خروجی پیش‌بینی می‌شود؛

خلاصه‌سازی؛ در این روش، بر اساس داده‌های ورودی خلاصه‌ای از معنای داده‌های ورودی استخراج می‌شود؛

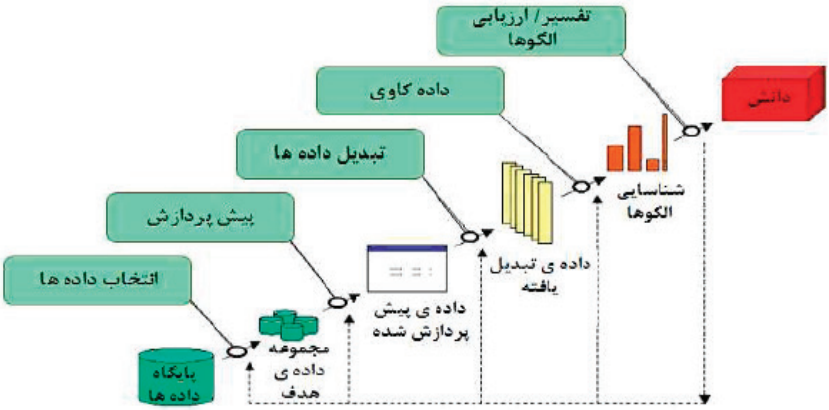
مدل‌سازی وابسته؛ در این روش، بر اساس داده‌ها و متغیرهای ورودی میزان وابستگی بین این متغیرها استخراج می‌شود؛

کشف تغییر و انحرافات؛ در این روش، بر اساس مقایسه داده‌های گذشته با حال، میزان تغییرات اعمال شده در داده‌های فعلی مشخص می‌شود.

به‌کارگیری الگوریتم‌های استخراج برای تحلیل داده‌ها و تولید تعداد خاصی از الگوها یا مدل‌هایی از داده‌ها، منظور از الگوریتم‌های داده‌کاوی ارائه مدل، ارزیابی مدل و روش جستجو است؛

تفسیر الگوهای حاصل از داده‌کاوی برای به دست آوردن دانش حاصل از داده‌ها (کرمی، ۱۳۸۶).

1. Knowledge Discovery in Database (KDD)



شکل ۷. نمایش اجمالی فرایند کشف دانش **(Fayyad, Piatetsky-Shapiro & Smyth, 1996)**

متن‌کاوی

از جمله مشکلاتی که در زمینه داده‌کاوی وجود دارد کشف دانش مفید از متون نیمه ساخت‌یافته یا غیرساخت‌یافته است که توجه زیادی را به خود جلب کرده است. روش‌های داده‌کاوی سنتی فرض بر این دارند که اطلاعات به فرم متنی پایگاه داده‌های رابطه‌ای هستند؛ به همین دلیل، برای بسیاری از کاربردها مانند اطلاعات الکترونیکی قابل دسترس به فرم متنی نیمه ساخت‌یافته یا غیرساخت‌یافته مفید نیستند. بدون عمل متن‌کاوی، پردازش پایگاه داده‌های متنی غیرساخت‌یافته باید به صورت دستی توسط کاربران انجام شود. بنابراین می‌توان گفت هدف متن‌کاوی خودکار کردن مقدار زیادی از عمل کاربران است.

متن‌کاوی[1] که به «تحلیل هوشمند متن»، «داده‌کاوی متنی» یا «کشف دانش در متن»، نیز مشهور است، به طور کلی به فرایند استخراج دانش و اطلاعات مورد علاقه و مهم از مجموعه متنی غیرساختاریافته اشاره دارد. به عبارت دیگر، متن‌کاوی فرایند تحلیل طبیعی متن به منظور کشف و ثبت اطلاعات معنایی برای درونداد و ذخیره‌سازی در یک ساختار سازمان‌دهی شده دانش است (Pons-Porrata, Berlanga-Llavori & Ruiz-Shulcloper, 2007)

نخستین بار «لان»[2] بود که در سال ۱۹۵۸، مفهوم متن‌کاوی را در مقاله خود مطرح کرد. در سال ۱۹۶۱ نیز «دویله»[3] در مقاله‌ای به متن‌کاوی و روش‌های مرتبط با آن پرداخت و بیان داشت که «طبقه‌بندی و سازماندهی اطلاعات» می‌تواند از تجزیه و تحلیل، تکرار و توزیع واژگان به‌کار رفته در آن صورت پذیرد. اگر این دو مورد را به عنوان نخستین موارد مطرح شدن متن‌کاوی در نظر بگیریم،

1. Knowledge Discovery and Text mining (KDT)
2. H. P. Luhn
3. Doyle

نتیجه‌گیری می‌شود که متن‌کاوی ممکن است مفهومی جدید باشد؛ اما رؤیای استخراج خودکار اطلاعات از متون، همزمان با پیدایش رایانه وجود داشته است (شارپ، ۲۰۰۱، به نقل از جبرئیلزاده، ۱۳۹۱).

سوانسون[1] (۱۹۸۸) در مقاله خود بیان داشت که متون علمی باید به عنوان پدیده‌های ارزشمند طبیعی، کشف، اصلاح و تجزیه و تحلیل شوند و به این ترتیب در واقع نظر دانشمندان را به استفاده از اطلاعات با تحلیل هوشمندانه جلب کرد. سوانسون با ایجاد یک نرم‌افزار، نخستین گام را در راستای عملی‌سازی متن‌کاوی طی کرد. او از این نرم‌افزار در استخراج اطلاعات مفید از متون پزشکی استفاده کرد (Swanson & Smalheiser, 1999). سوانسون هیچ‌گاه از اصطلاح متن‌کاوی برای نرم‌افزار خود استفاده نکرد، اما به نظر می‌رسد که این نرم‌افزار نخستین نرم‌افزار متن‌کاوی بوده است. از این رو می‌توان او را پدر متن‌کاوی مدرن نامید. لیندزی، گوردن و کاتساف[2] (۱۹۹۹) کارهای سوانسون را – بدون آنکه نام متن‌کاوی بر آن اطلاق کنند– ادامه دادند. نرم‌افزار آنها دو واژه را به صورت هم‌زمان در میان متون جستجو می‌کرد و نتیجه را در فهرستی قرار می‌داد تا کاربران آنها را به عنوان مقاله‌های تکمیلی مورد مطالعه قرار دهند. لیندزی و گوردن در همان سال روش TF-IDF را برای رتبه‌بندی متون و واژه‌ها به این نرم‌افزار افزودند (اسمال‌هیزر و گوردن، ۱۹۹۹، به‌نقل از جبرئیل‌زاده، ۱۳۹۱). اما نخستین محصول نرم‌افزاری در زمینه متن‌کاوی، در سال ۱۹۹۹ باعنوان «استخراج هوشمند متن»[3] ، توسط شرکت «آی بی ام»[4] به بازار عرضه شد (Dörre, Gerstl & Seiffert , 1999).

با توسعه مفهوم متن‌کاوی، مفاهیم دیگری نیز پابه‌پای آن رشد کردند؛ «بازیابی اطلاعات از مجموعه متون»[5]، «بازیابی اطلاعات از یک متن»[6]، «کشف دانش از بانک‌های اطلاعاتی»[7]، «مدیریت دانش در سازمان‌ها»[8] و «نمایش (مصورسازی) داده‌ها و اطلاعات»[9]. این مفاهیم توسط کاتساف، اراد[10] و لوزیویچ[11] در سال ۲۰۰۰ در چند مقاله منتشر و سعی شد تا میان این مفاهیم تمایز قایل شوند. در سال ۲۰۰۱ میلادی کاتساف و دمارکو[12] متن‌کاوی را با «استخراج اطلاعات از متون فنی» تعریف

1. Swanson
2. Lindsay, Gordon & Kotsoff
3. Intelligent Miner for Text
4. IBM
5. Information Retrieval from Text Collections
6. Information Extraction from Individual Texts
7. Knowledge Discovery in Databases
8. Knowledge Management in Organizations
9.Visualization of Data and Information
10. Orad
11. Losiewicz
12.DeMarco

کردند. بر اساس این تعریف، متن‌کاوی شامل سه بخش می‌شود؛ «بازیابی اطلاعات»، «پردازش اطلاعات» و «یکپارچگی اطلاعات»[1]. پردازش اطلاعات به استخراج الگوهای موجود در متون بازیابی شده اطلاق می‌شود، و یکپارچگی اطلاعات، ترکیب هم‌افزایانه خروجی رایانه‌ای پردازش اطلاعات با خواندن اطلاعات بازیابی شده توسط انسان است، که مفهوم سیستم انسان– ماشین را تداعی می‌کند (Kostoff & Demarco, 2001).

متن‌کاوی تکیه‌اش روی پیدا کردن دانش[2] جدید از متن است (معمولاً دانشی که به طور ضمنی در سندهای متنی است) در حالی که بازیابی اطلاعات، سندهایی می‌یابد که بیشترین ارتباط را دارند. هدف اولیه متن‌کاوی، بازیابی اطلاعات از متون ساخت‌نیافته و همچنین ارائه دانش به صورت خالص برای کاربران در شکل چکیده است (Ananiadou, S. & McNaught, 2006). متن‌کاوی، قادر ساختن استفاده‌کنندگان برای جمع‌آوری، ذخیره، تفسیر و کشف دانش مورد نیاز برای تحقیق و آموزش مؤثر و نظام‌مند است. در کل می‌توان بیان داشت، هدف متن‌کاوی، کشف اطلاعات از قبل ناشناخته است که هنوز کسی نمی‌داند و بنابراین مستند نشده است(Dörre, Gerstl & Seiffert , 1999). متن‌کاوی شامل سه فعالیت بزرگ است؛ «بازیابی اطلاعات»، که بازیابی متون مربوط به سؤال استفاده‌کنندگان است؛ «خلاصه اطلاعات»، که شناختن و استخراج نکات ریز متون که مربوط به سؤال هستند؛ و «داده‌کاوی»، که رابطه مستقیم یا غیرمستقیم بین قسمت‌های اطلاعات استخراجی از متون را پیدا می‌کند(Thomas, McNaught & Ananiadou, 2011).

از نظر فایاد[3] و همکارانش (۱۹۹۶)، کشف دانش، فرایند غیربدیهی تشخیص الگوهای معتبر، نو، مفید و در نهایت قابل درک در داده‌ها است. فرق داده‌کاوی با متن‌کاوی این است که در متن‌کاوی الگوها از متن زبان طبیعی استخراج می‌شوند، در حالی که در داده‌کاوی الگوها را از پایگاه‌های داده ساخت‌یافته به دست می‌آورند. متن‌کاوی، متصل کردن اطلاعات استخراج شده به یکدیگر برای تشکیل حقایق یا فرضیه‌های جدید است تا پس از آن به کمک روش‌های متعارف آزمایشی، بررسی بیشتری شوند.

متن‌کاوی درباره جستجوی انگاره‌ها در متن زبان طبیعی است. آن عبارت است از، فرایند تحلیل متن به منظور استخراج اطلاعات از حجم عظیمی از متون غیرساختاریافته آزاد. به عبارتی، متن‌کاوی فناوری است که امکان کشف انگاره‌ها و گرایش‌ها را به طور نسبتاً خودکار از درون متن غیرساختاریافته آزاد فراهم می‌کند. همانطوری که بیان شد، متن‌کاوی به عنوان امتداد طبیعی داده‌کاوی به شمار می‌رود. متن‌کاوی در مقایسه با داده‌کاوی فرایند پیچیده‌تری است، زیرا با آن دسته از داده‌های متنی ارتباط دارد که ذاتاً غیرساختاریافته هستند. متن‌کاوی حوزه‌ای چند رشته‌ای است و شامل بازیابی اطلاعات، تحلیل متن، دسته‌بندی اطلاعات، طبقه‌بندی اطلاعات، مصورسازی اطلاعات، فناوری پایگاه‌های

1. Information Integration
2. knowledge
3. Fayyad

اطلاعاتی، یادگیری ماشینی و داده‌کاوی است. علاوه بر این، از ابزارهای متن‌کاوی برای کمک به پژوهشگران در زمینه‌های مختلف استفاده می‌شود. به عنوان مثال، یک ستاره‌شناس که علائم اشعه ایکس را در ناحیه‌ای از فضا کشف می‌کند ممکن است تمایل داشته باشد که در پیشینه‌های پیوسته[1] جستجو کند تا ببیند که آیا تاکنون هیچ نوع اشعه مادون قرمزی در همان ناحیه کشف شده است یا خیر. یک زیست‌شناس که دارای فهرستی از ۱۰۰ ژن است که در مقاله‌های مختلف شناسایی شده‌اند، ممکن است تمایل داشته باشد که به سراغ تحقیقات منتشر شده موجود برود و در جستجوی مقاله‌هایی باشد که درباره عملکرد این ژن‌ها توضیح داده باشند (Sharma, 2005).

هیرست[2] نیز تعریف نسبتاً دقیقی از متن‌کاوی ارائه داده و با تعریف خود، آن را از دسترسی به اطلاعات (بازیابی سنتی اطلاعات) متمایز ساخته است. بازیابی سنتی اطلاعات، بیشتر روی بازیابی متون مرتبط با هم تأکید می‌کند، متونی که با نیازهای اطلاعاتی کاربر مرتبط باشد. بر اساس تعریف هیرست، داده‌کاوی تنها با اطلاعات و بازیابی آن ارتباط ندارد، بلکه تلاش می‌کند تا اطلاعات جدید را از میان داده‌ها کشف کند که پیش از این- حتی برای ایجادکننده داده‌ها (نویسنده متن) هم- مشخص نبود. او معتقد بود که داده‌کاوی و متن‌کاوی «توأم با شانس»[3] هستند. درحالی‌که بازیابی اطلاعات «هدف‌گرا»[4] است. بنابراین کارهایی نظیر جستجوی واژه‌ها برای پاسخ به پرسش‌ها، متن‌کاوی به شمار نمی‌رود. او بر این باور بود که بازیابی و دستیابی به اطلاعات می‌تواند به عنوان کار تکمیلی و پشتیبان برای متن‌کاوی به شمار رود (Hearst, 1997).

متن‌کاوی کاربردهای متعددی دارد؛ از جمله، بررسی روندهای علمی و خوشه‌بندی و دسته‌بندی مفهومی متون علمی و صفحه‌های وبی و غیره. دوروری و همکارانش بر این باورند، از آنجا که حدود ۸۵ تا ۹۰ درصد از داده‌های شرکت‌ها و سازمان‌ها با روش‌های متعارف داده‌کاوی قابل کشف دانشی نیستند و از اسناد متنی تشکیل یافته‌اند؛ چه بسا متن‌کاوی وظیفه پیچیده‌تری نسبت به داده‌کاوی سنتی دارد. چون مستلزم برخورد کردن با اطلاعات متنی غیرساختاریافته است که ذاتاً مبهم هستند (Dörre, Gerstl & Seiffert , 1999; Hotho, Nürnberger & Paaß, 2005).

نتیجه‌ای که از تعاریف فوق به دست می‌آید، اینکه «متن‌کاوی»، شاخه‌ای از علم داده‌کاوی است که برای کشف دانش ضمنی در فایل‌های متنی غیرساختارمند، به صورت خودکار و با استفاده از فنونِ وام گرفته‌شده از دانش داده‌کاوی عمل می‌کند.

1. Online
2. Hearst
3. Opportunistic
4. Goal- Driven

روش‌های متن‌کاوی

روش‌های متعددی برای متن‌کاوی و ابزارهای وابسته به آنها وجود دارد؛ شناخت خودکار واژه[1]، خوشه‌بندی مدارک[2] ، طبقه‌بندی (دسته‌بندی) مدارک[3]، خلاصه‌سازی مدارک[4]، که در زیر توضیح مختصری راجع به هر یک از آنها ارائه شده است:

شناخت خودکار واژه: هر مرور نظام‌مندی به ما تمرکز روی یک موضوع یا موضوع‌های بیشتر را می‌دهد که می‌توانند در یک شبکه مفهوم و ارتباط بین آنها جمع شود. این مفاهیم در متون به عنوان واژه‌های فنی (کلیدواژه) شناخته شده‌اند، که بر خلاف واژه‌های عمومی زبان، به عنوان هدف اصلی برای طبقه‌بندی دانش هستند. شناخت خودکار واژه، فنی است که به صورت خودکار واژه‌ها را از متون می‌شناسد و استخراج می‌کند(Ananiadou & McNaught, 2006). اینها به نوعی با مفهوم دامنه‌ای برابرند که برای اصطلاح‌نامه، واژگان کنترل‌شده و هستی‌شناسی[5] است. اگرچه واژه‌های استخراجی اغلب می‌توانند مثل واژه‌های نمایه یا کلیدواژه‌ها باشند که توسط نویسنده داده می‌شود، آنها نیز از متن به صورت خودکار اختصاص داده می‌شوند(Thomas, Mc-Naught & Ananiadou, 2011). روش‌های متفاوتی از روش شناخت خودکار واژه وجود دارد که عبارتند از: «پایه نقش»[6]، «آماری»[7]، «هم‌رخدادی»[8] و «فنون یادگیری ماشینی»[9]. روش‌های شناخت خودکار واژه شامل شناخت حدود واژه‌های چندکلمه‌ای یا واژه‌های متداخل (برای مثال Systematic Review Process دو عبارتSystematic review و Review Process را شامل می‌شود) است.

خوشه‌بندی مدرک: خوشه‌بندی روشی است که برای گروه‌بندی موجودیت‌های[10] (مدارک) مشابه مورد استفاده قرار می‌گیرد. در این روش، مدارک در گروه‌هایی از پیش تعیین نشده به نام خوشه قرار می‌گیرند؛ به طوری که مدارک مشابه در کنار یکدیگر و مدارک نامشابه دور از یکدیگر قرار می‌گیرند(Tan, Steinbach & Kumar, 2006). به این ترتیب می‌توان بیان داشت که، ورودی اصلی یک الگوریتم خوشه‌بندی، اندازه فاصله[11] است. هدف خوشه‌بندی مفهومی مدرک، گروه‌بندی

1. Automatic Term Recognition (ART)
2. Document Clustering
3.Document Classification
4. Document Summarization
5. Anthology
6. Rule- Based
7.Statistical
8. Co-Occurrence-Based
9. Machine Learning Techniques
10. Entity
11. Distance Measure

مجموعه‌ای از مدارک بر پایه موضوع آنها است. گروه‌هایی که ایجاد می‌شوند به عنوان خوشه‌های موضوعی دیده می‌شوند، گرچه شناخت موضوع یک مدرک یک فرایند ساده برای فناوری‌های متن‌کاوی نیست. از نگرش‌های رایج برای خوشه‌بندی مدارک، واژه‌های مدارک (نگرش جعبه واژگان)[1] و تعداد رخداد آنها در یک مدرک برای شناخت موضوع هسته است. خوشه‌بندی آثار بهتر است که با مجموعه مدارک بزرگ انجام شود، چون نویزها کاهش می‌یابد و اجازه می‌دهد که با دید کامل‌تری خوشه‌بندی انجام گیرد (Thomas, McNaught & Ananiadou, 2011).

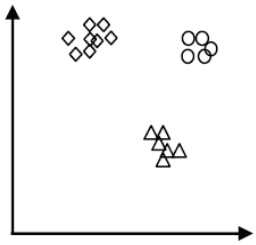
در انواع خوشه‌بندی می‌توان آنها را به دو دسته زیر تقسیم کرد:

«**خوشه‌بندی مسطح**»[2] (شکل ۹)، که مجموعه‌ای مسطح از خوشه‌ها را بدون ساختار روشنی ارائه می‌دهد که خوشه‌ها را به همدیگر مرتبط سازد؛

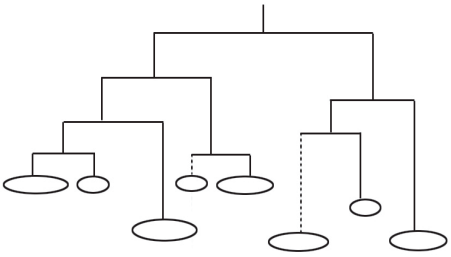
«**خوشه‌بندی سلسله مراتبی**»[3] (شکل ۱۰)، که خوشه‌ها را در ساختاری سلسله مراتبی ارائه می‌دهد.

دومین تمایز مهم بین الگوریتم‌های خوشه‌بندی سخت و نرم برقرار است. «خوشه‌بندی سخت»[4] یک تقسیم‌بندی بدون انعطاف از اسناد ارائه می‌دهد، یعنی هر سند می‌تواند فقط عضو یک خوشه باشد. تخصیص الگوریتم «خوشه‌بندی نرم»[5] انعطاف‌پذیر است، یعنی تخصیص هر سند، توزیعی روی تمامی خوشه‌هاست. در این تخصیص، هر سند در چند خوشه عضویت دارد، اما رتبه عضویت آن ممکن است در هر خوشه متفاوت باشد. «نمایه‌سازی معنایی پنهان»[6] نمونه‌ای از الگوریتم خوشه‌بندی نرم است. خوشه‌بندی سخت دارای ساختار مسطح، هر سند فقط عضو یک خوشه است، اما در ساختار سلسله مراتبی، امکان عضویت کامل بیش از یک خوشه وجود دارد. به عبارت دیگر، در «خوشه‌بندی چند سطحیِ»[7] سلسله مراتبی تمامی اعضای یک خوشه عضو خوشه لایه بالاتر از خود هستند. تفاوت خوشه‌بندی سخت که امکان عضویت چندگانه را فراهم می‌کند با خوشه‌بندی نرم این است که در خوشه‌بندی سخت مقادیر یا صفر است یا ۱، در حالی که در خوشه‌بندی نرم می‌توانند غیر از عدد منفی هر مقداری را داشته باشند (حسینی، ۱۳۹۰).

1. Bag- of- Words Approach
2. Flat Clustering
3. Hierarchical Clustering
4. Hard Clustering
5. Soft Clustering
6. Latent Semantic Indexing
7. Partitional Clustering



شکل ۹. مجموعه‌ای از داده‌ها دارای ساختار خوشه‌بندی مسطح (روشن)



شکل ۱۰. ساختار خوشه‌بندی سلسله‌مراتبی

طبقه‌بندی (دسته‌بندی) خودکار مدارک: طبقه‌بندی مدارک در متون مختلف با اصطلاح‌های مختلف از جمله Text Categorization، Text Classification و Document Categorization ذکر شده است. طبقه‌بندی مدارک در فرایند گروه‌بندی مدارک شبیه به هم در مجاورت یکدیگر است، که به موجب آن دانش به وسیله حذف داده‌های غیر ضروری، به‌صورت دقیق سازمان‌دهی می‌شود. طبقه‌بندی مدارک داخل مجموع‌های از مقوله‌های از پیش تعیین شده موجب بازیابی، سازمان‌دهی، تصویرسازی، توسعه و انتقال دانش به صورتی کارآمد می‌شود. در کل دو شکل تجزیه و تحلیل در طبقه‌بندی مدارک وجود دارد: دسته‌بندی مدارک به عنوان یک نگرش نظارتی و خوشه‌بندی مدارک به عنوان یک نگرش بدون نظارت (Ananiadou et al., 2009).

دسته‌بندی، هر جزء از داده‌ها را بر مبنای اختلاف بین داده‌ها به مجموعه‌های از پیش تعریف شده دسته‌ها تصور می‌کند. در حالی که خوشه‌بندی، داده‌ها را به گروه‌های مختلف (خوشه‌ها) که از قبل معین نیستند، (بر اساس مشابهت درون خوشه و تفاوت بیرون خوشه) تقسیم می‌کند. بنابراین اگر بخواهیم با استفاده از مفهوم یادگیری، دسته‌بندی و خوشه‌بندی را متمایز کنیم، باید بگوئیم دسته‌بندی، یادگیری با نظارت و خوشه‌بندی، یادگیری بدون نظارت است. یادگیری با نظارت یا دسته‌بندی عبارت است از، یادگیری به وسیله نمونه‌ها. به بیان دیگر، در این روش دسته‌ها از قبل مشخص هستند؛ ولی در یادگیری بدون نظارت یا خوشه‌بندی، خوشه‌ها مشخص نیستند و هدف خوشه‌بندی، تعیین خوشه‌های داده‌ها است (غضنفری، علیزاده و تیمورپور، ۱۳۸۷). طبقه‌بندی خودکار مدارک، در مراحل

بعدی می‌تواند به جستجو و جدا کردن در مرورهای خودکار، کاهش زمان بین شناخت و جداسازی متون مربوط به یک سؤال خاص کمک کند. با رشد متون، به فنون خودکار که مرورگر را در جداسازی مواردی که مورد نیاز هستند کمک کنند، مورد نیاز است (Ananiadou et al., 2009).

طبقه‌بندی متن به فرایند تخصیص خودکار یک یا چند سند به یکی از دسته‌های موضوعی است، که قبلاً تعریف شده است. در واقع طبقه‌بندی متن یکی از روش‌های شناسایی الگو برای استفاده مؤثر از اسناد است. این زمینه کاری در قسمت مهندسی امروزه به صورت وسیع مورد پژوهش و بررسی قرار گرفته است. در بخش صنعت هم، سیستم‌های طبقه‌بندی، به صورت جداگانه و مستقل برای پشتیبانی سیستم‌های اطلاعاتی مورد استفاده قرار گرفته‌اند (Thomas, McNaught & Ananiadou, 2011).

خلاصه‌سازی مدارک: خلاصه‌سازی مدارک و استخراج جمله‌های برجسته، روی این فرضیه پایه‌گذاری شده است که خواننده هر جمله را به مجموعه‌ای از اطلاعات خردشده می‌شکند که جمله‌ها ارائه می‌دهند. اطلاعات خردشده متقابلاً مستقل هستند و اطلاعات خردشده دارای یک مقیاس ارزشی است. چند روش وجود دارد که برای اصلاح کیفیت خلاصه‌سازی استفاده می‌شود؛ «استخراج جمله‌ها»، «ساخت جمله»، «حذف جمله‌های زائد»، «مرتب‌سازی جمله‌ها»، «آشکارسازی موضوع» و «استخراج اطلاعات»(Thomas, McNaught & Ananiadou, 2011).

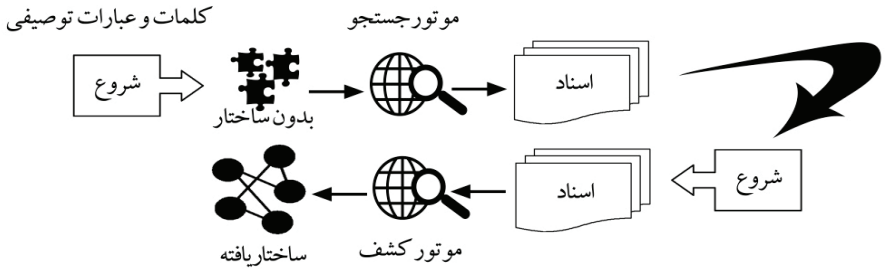
فرایند متن‌کاوی

یکی از دلایل استفاده از روش‌های داده‌کاوی روی اسناد متنی ایجاد ساختار در آنها است. ایجاد ساختار می‌تواند دسترسی کاربر را به یک مجموعه از اسناد به طور چشمگیری ساده کند. از ساختارهای دسترسی معروف می‌توان به کتاب فهرست‌های کتابخانه‌ای و نمایه‌های کتاب اشاره کرد. اما مسئله نمایه‌های طراحی شده به صورت دستی مسئله‌ای زمان‌بر است و این نمایه‌ها معمولاً به‌روز نیستند و برای انتشارات جدید و یا اطلاعات زودبه‌زود در حال تغییر هستند، مانند اطلاعات روی اینترنت، قابل استفاده نیستند. در واقع روش متن‌کاوی اغلب همان روش‌های داده‌کاوی هستند که با تغییراتی برای متون استفاده می‌شوند (Kontardzic, 2003). که در ادامه این روش تشریح شده است.

در شیوه‌های سنتی پردازش اطلاعات غیرساختاریافته، فرد یا افرادی در نظر گرفته می‌شدند تا با مطالعه داده‌های متنی، به تحلیل آنها بپردازند. متن‌کاوی، فرایندی است که به کمک آن اطلاعات ارزشمند و مفید نهفته در داده‌های غیرساختاریافته، استخراج می‌شود. در متن‌کاوی تحلیل محتوا داده‌های غیرساختاریافته بر اساس آنالیز کمّی متن صورت می‌پذیرد. چنانچه اطلاعات کمّی با سیستم‌های هوشمند مورد بررسی و تحلیل قرار گیرد، می‌توان تا اندازه‌ای به بررسی کیفی متون پرداخت و یک سیستم Interactive (Man- Machine) ایجاد کرد (Hearst, 1999).

فرایند متن‌کاوی (شکل ۱۱) شامل مراحل زیر است:

۱. **آماده‌سازی داده‌های متنی؛**
۲. **نادیده گرفتن واژگان بدون ارزش اطلاعاتی[1]؛**
۳. **ریشه‌یابی کلمات[2]؛**
۴. **جستجوی واژگان مهم و کلیدی متن؛**
۵. **آنالیز و تحلیل متن‌های مختلف بر اساس واژگان کلیدی** (همانند یافتن متن‌های مشابه، خوشه‌بندی کردن متن‌ها، تعیین موضوع متون و...) که از آن می‌توان به عنوان لایه کاربردی فرایند متن‌کاوی نام برد.



شکل ۱۱. فرایند متن‌کاوی (کرمی، ۱۳۸۶)

در آماده‌سازی داده‌های متنی، چگونگی ذخیره داده‌های غیرساخت‌یافته، نمایه‌سازی آنها، تبدیل فرمت آنها به فرمت‌های قابل استفاده در نرم‌افزارها مورد بررسی قرار می‌گیرد. نادیده گرفتن واژگان بدون محتوای با ارزش اطلاعاتی همانند: از، به، با، تا و ... (که معمولاً بیشترین میزان تکرار در متن را به خود اختصاص می‌دهند) در مرحله دوم مورد بررسی قرار می‌گیرد. از آنجا که در متن‌کاوی به دنبال واژگان ارزشمند اطلاعاتی هستیم، این دسته از واژگان را کنار خواهیم گذاشت. ریشه‌یابی واژگان در مرحله سوم مورد بررسی قرار می‌گیرد. واژگانی مانند "دانش"، "دانشمند"، "دانائی" و ... واژگان هم‌ریشه به حساب می‌آیند. در بررسی یک متن توسط فرایند متن‌کاوی به تعیین ریشه و واژگانی هم‌ریشه پرداخته می‌شود تا در بررسی آماری واژگان متن، این دسته از واژگان در کنار هم مورد بررسی قرار گیرد. تعیین واژگان کلیدی و مهم متن مهم‌ترین بخش فرایند متن‌کاوی است و دقت تمامی کاربردهای متن‌کاوی، تا حدود زیادی به دقت در تعیین واژگان کلیدی بستگی خواهد داشت. برای تعیین واژگان کلیدی و مهم متن الگوریتم‌های مختلفی بر اساس تکنیک‌های ریاضی ارائه شده است. روش IDF–TF ، روش تکرار واژگان، روش تکرار واژگان معمولی شده، روش وزن‌دهی بر اساس جستجوی کاربر و سایر الگوریتم‌های مشتق شده از این الگوریتم‌ها در تعیین واژگان کلیدی و مهم به کار گرفته

1.Stop Word
2. Stemming

می‌شوند. وجه اشتراک تمامی این الگوریتم‌ها آن است که تعیین واژگان کلیدی یک متن بر اساس آنالیز یک مجموعه متون مرتبط با هم تعیین می‌شود. کاربردهای مختلفی را می‌توان از فرایند متن‌کاوی انتظار داشت. بر اساس واژگان کلیدی می‌توان با الگوریتم‌های خوشه‌بندی«کا- مینز»، «سلسله‌مراتبی» و یا سایر الگوریتم‌ها به خوشه‌بندی متون پرداخت. با استفاده از خوشه‌بندی متون می‌توان متن‌های مشابه را شناسایی و دسته‌بندی کرد. با تعیین واژگان کلیدی و استفاده از تکنیک‌های مختلف وزن‌دهی به جملات و واژگان متن می‌توان یک چکیده از متن تهیه و ارائه کرد (زعفریان، ۱۳۸۵).

نتیجه‌گیری

مصورسازی دانش یا بازنمون دانش و در عرصه‌های محدودتر مصورسازی یک حوزه علمی از طریق مصورسازی ساختار درونی آن، به کاربران کمک خواهد کرد تا به سرعت درک شفافی از ساختار حوزه مورد نظر با مشاهده گره‌ها، پیوندها و فاصله‌ها (به عنوان نمادهای گرافیکی) داشته باشند. در واقع می‌توان اینگونه بیان داشت، که نقشه‌های علمی واقعیت‌های علم هستند و خلاءها و حوزه‌های داغ فعالیت‌های علمی را ترسیم می‌کنند، و به عبارت دیگر، نقشه راهی برای آینده‌نگری دانشمندان و سیاست‌گذاران حوزه‌های مختلف علوم هستند.

همانگونه که سخن رفت، در بحث از نقشه‌های علم دو رویکرد مورد توجه است: نخست، نقشه به عنوان آنچه تاکنون در حوزه موضوعی خاص صورت گرفته و دوم، نقشه به عنوان مسیر تحقیقات آینده در رویکرد اول، با توجه به مقاله‌ها و مجله‌ها و کارهای صورت گرفته در گذشته و در یک بازه زمانی خاص به ترسیم نقشه پرداخته می‌شود و آنچه تاکنون صورت گرفته مورد بررسی و بحث قرار می‌گیرد؛ و در رویکرد دوم، هدف از ترسیم نقشه‌های علم، در واقع راهنمایی برای انجام پژوهش‌های آینده است. البته این دو رویکرد کاملاً نسبت به هم بیگانه نیستند، با ترسیم نقشه‌های علم بر اساس بازه‌های زمانی گذشته می‌توان به قوت‌ها و ضعف‌های پژوهش‌های انجام گرفته، هم از حیث حجم کارهای صورت گرفته و هم از لحاظ حوزه‌هایی که مورد غفلت واقع شده و یا بیش از حد به آنها پرداخته شده، آگاه شد و برای اصلاح روند موجود، در نقشه‌هایی که به ترسیم وضع مطلوب در آینده می‌پردازند، راهکارهای مناسب در نظر گرفت (عابدی جعفری و همکاران، ۱۳۹۰، ۵۹).

نوروزی چاکلی (۱۳۹۰) به درستی تصریح کرده است و در مورد اهمیت مصورسازی علم و تأثیر شگرفی که می‌تواند بر درک بهتر روابط موجود در علم داشته باشد سخن گفته است؛ و بر اساس این فرض، «دانش ضمنی» را که در گذشته برای بیشتر دانشمندان حوزه‌های مختلف درک شدنی بود، به صورت آشکار می‌توان نمایش داد و آن را به «دانش آشکار» تبدیل کرد. برای این مهم، دانشمندان عرصه علم‌سنجی و تصویرنگاری علوم به طور معمول از روش‌های «تحلیل هم‌استنادی» و یا روش «تحلیل هم‌رخدادی واژگان» بهره می‌گیرند. اما امروزه پرکاربردترینِ روش‌ها، تحلیل

هم‌رخدادی واژگان علوم است. علت اینکه تمایلات به تحلیل‌های هم‌رخدادی نسبت به تحلیل‌های هم‌استنادی بیشتر شده است، را می‌توان دو چیز دانست؛ از نظر علمی، در روش هم‌رخدادی واژگان، می‌توان مدارکی که به آنها استناد نشده است را نیز در تحلیل وارد کرد. همچنین از نظر روش، در تحلیل هم‌استنادی، تحلیل نحوه پویایی (دینامیک) حوزه مورد پژوهش با روندهای موجود در عمل نویسندگان مقاله‌ها، آمیخته می‌شود و برآورد خالصی از پیشرفت علم به ما نمی‌دهد. در روش‌های تحلیل هم‌رخدادی، ماتریس‌های مفهومی شکل می‌یابند، روش‌هایی که از فنون متن‌کاوی وام گرفته شده است. بیشترین فنونی که در تحلیل‌های متن‌کاوی مورد استعمال واقع می‌شوند، می‌توان به شناخت خودکار واژه‌ها، خوشه‌بندی مدارک، طبقه‌بندی (دسته‌بندی) مدارک و خلاصه کردن مدارک اشاره کرد. امروزه در تحقیقات پژوهشگران علم‌سنجی ایرانی، روش‌های خوشه‌بندی و دسته‌بندی جایگاه ویژه‌ای در ترسیم نقشه‌های علمی و مصورسازی اطلاعات علمی یافته‌اند، و سیر بلوغی خود را با تکامل نرم‌افزارهای مصورسازی طی می‌کنند. در این نوشته تلاش بر این بود تا نگاهی گسترده‌تر به فنون و روش‌های علم‌سنجی داشته باشیم، امید است پژوهشگران حوزه علم‌سنجی کشورمان با این دانش جدید در مصورسازی ساختار علوم آشنا شوند و در تحقیقات علمی خود از آنها بهره گیرند.

منابع

پشوتنی‌زاده، م.؛ عصاره، ف. (۱۳۸۸). تحلیل استنادی و ترسیم نقشه تاریخ‌نگاشتی تولیدات علمی کشاورزی در نمایه استنادی علوم در سال‌های ۲۰۰۰ تا ۲۰۰۸. *فصلنامه علوم و فناوری اطلاعات*، *۲۵* (۱)، ۲۳-۵۲.

جبرئیل‌زاده، ح. (۱۳۹۱). *نگاشت نقشه علمی کتابداری بر اساس پایان‌نامه‌های رشته کتابداری ایران*. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، تهران، دانشگاه علامه طباطبائی .

حرّی، ع. (۱۳۸۱). تحلیل استنادی. در *دایرةالمعارف کتابداری و اطلاع‌رسانی*. (ج. ۱، ص ۶۱۶- ۶۲۰). تهران: کتابخانه ملی جمهوری اسلامی ایران.

حسینی، س. م. (۱۳۹۰). بررسی عناصر و مؤلفه‌های رابط کاربر در نظام‌های بازیابی اطلاعات مبتنی بر خوشه‌بندی. *فصلنامه علوم و فناوری اطلاعات*، *۲۶* (۳)، ۶۲۵-۶۵۳.

حیدری، غ. (۱۳۸۹). *معرفت‌شناسی علم‌سنجی*. شیراز: نوید شیراز.

درودی، ف. (۱۳۸۷). مبانی و راهبردهای ارائه و نمایش دیداری اطلاعات. *فصلنامه علوم و فناوری اطلاعات*، *۲۳* (۴)، ۱۰۳-۱۲۶.

رحیمی، م.؛ فتاحی، ر. (۱۳۸۶). همکاری علمی و تولید اطلاعات: نگاهی به مفاهیم و الگوهای رایج در تولید علمی مشترک. *فصلنامه کتاب*، *۱۸* (۳)، ۲۳۵-۲۴۸.

زعفریان، ر. (۱۳۸۵). *روش‌هایی برای متن‌کاوی متون فارسی به همراه مطالعه موردی در مهندسی صنایع*. پایان‌نامه دکتری، تهران، دانشگاه صنعتی شریف.

زوارقی، ر.؛ فدایی، غ.؛ فهیم‌نیا، ف. (۱۳۹۰). چشم‌اندازی بر مبانی نظری مصورسازی حوزه دانش. *فصلنامه*

تحقیقات کتابداری و اطلاع‌رسانی دانشگاهی، ۴۵(۵۷)، ۱۳-۳۷.
عابدی جعفری، ح.؛ ابویی اردکان، م.؛ آقازاده دهده، ف.؛ دلبری راغب، ف. (۱۳۹۰). روش‌شناسی ترسیم نقشه‌های علم: مطالعه موردی ترسیم نقشه علم مدیریت دولتی. *روش‌شناسی علوم انسانی، ۱۷* (۶۶)، ۵۳-۶۹.
عصاره، ف.؛ حیدری، غ.؛ زارع فراشبندی، ف.؛ حاجی زین‌العابدینی، م. (۱۳۸۸). *از کتاب‌سنجی تا وب‌سنجی: تحلیلی بر مبانی، دیدگاه‌ها، قواعد و شاخص‌ها*. با مقدمه عباس حری. تهران: کتابدار.
عطاپور، ه.؛ حسن‌زاده، م.؛ نوروزی چاکلی، ع. (۱۳۸۸). رابطه خوداستنادی و هم‌آیندی مؤلفان با ضریب تأثیر نشریات در ایران: مطالعه موردی نشریات حوزه اقتصاد. *فصلنامه علوم و فناوری اطلاعات، ۲۵* (۲)، ۲۰۷-۲۲۶.
غضنفری، م.؛ علیزاده، س.؛ تیمورپور، ب. (۱۳۸۷). *داده‌کاوی و کشف دانش*. تهران: دانشگاه علم و صنعت ایران.
کانتاردزیک، م. (۱۳۸۵). *داده‌کاوی* (امیر علیخانزاده، مترجم). بابل: علوم رایانه.
کرمی، م. (۱۳۸۶). کاربرد ابزارهای تحلیلگر داده‌کاوی و متن‌کاوی در چابکی سازمان‌های مراقبت بهداشتی و درمانی. *فصلنامه مدیریت سلامت، ۱۰* (۳۰)، ۱۵-۲۰.
محمدی، ا. (۱۳۸۷). *ترسیم نقشه علمی نانوتکنولوژی در ایران*. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، تهران: دانشگاه آزاد اسلامی، واحد علوم و تحقیقات.
مدیر امانی، پ. (۱۳۸۱). استناد. در *دایره‌المعارف کتابداری و اطلاع‌رسانی*. (ج ۱، ص ۱۷۹). تهران: کتابخانه ملی جمهوری اسلامی ایران.
ناصری جزه، م.؛ طباطبائیان، ح.؛ فاتح راد، م. (۱۳۹۱). ترسیم نقشه دانش مدیریت فناوری در ایران با هدف کمک به سیاست‌گذاری دانش در این حوزه. *فصلنامه سیاست علم و فناوری، ۵*(۱)، ۴۵-۷۲.
نوروزی چاکلی، ع. (۱۳۹۰). *آشنایی با علم‌سنجی (مبانی، مفاهیم، روابط و ریشه‌ها)*. تهران: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت)، مرکز تحقیق و توسعه علوم انسانی؛ دانشگاه شاهد، مرکز چاپ و انتشارات، ۱۳۹۰.
Agutter, J., & Bermudez, J. (2005). *Information visualization design: The growing challenges of a data saturated world*. AIA Report on university research, 61-75. Retrieved September 17, 2013, from http://www.aia.org/SiteObjects/files/Agutter_ color.pdf
Ananiadou, S., & McNaught, J. (2006). *Text mining for biology and biomedicine*. Boston, London: Artech House.
Ananiadou, S., Rea, B., Okazaki, N., Procter, R., & Thomas, J. (2009). Supporting systematic reviews using text mining. *Social Science Computer Review, 27*(4), 509-523.
Börner, K., & Scharnhorst, A. (2009). Visual conceptualizations and models

of science. *Journal of Informetrics, 3*(3), 161-172.

Börner, K., Chen, C., & Boyack, K. (2003). Visualizing Knowledge Domains. In Blaise Cronin (Ed.), *Annual Review of Information Science & Technology, Volume 37*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, chapter 5, pp. 179-255.

Boyack, K. W. (2004). Mapping knowledge domains: Characterizing PNAS. *Proceedings of the National Academy of Sciences, 101*(suppl 1), 5192-5199.

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., Law, J., & Rip, A. (1986). *Mapping the Dynamics of Science and Technology*. London: Macmillan.

Cheong, F., & Corbit, B. (2009, June 8-10). *A social network analysis of the co-authorship network of the Australian conferences of Information Systems from 1990 to 2006*. Paper Presented in the 17th European Conference on Information Systems, Verona, Italy.

Cronin, B., Shaw, D., & La Barre, K. (2004). Visible, less visible, and invisible work: Patterns of collaboration in 20th century chemistry. *Journal of the American Society for Information Science and Technology, 55*(2), 160-168.

Dörre, J., Gerstl, P., & Seiffert, R. (1999, August). Text mining: finding nuggets in mountains of textual data. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 398-401). ACM.

Duan, C. H. (2011). Mapping the intellectual structure of modern technology management. *Technology Analysis & Strategic Management, 23*(5), 583-600.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine, 17*(3), 37.

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Francisco, USA: Morgan Kaufman Publishers.

Harsanyi, M. A. (1993). Multiple authors, multiple problems bibliometrics

and the study of scholarly collaboration: A literature review. *Library and Information Science Review, 15*, 325–354.

He, Q. (1999). Knowledge Discovery Through Co-Word Analysis. *Library Trends, 48*(1), 133-159.

Hearst, M. A. (1997). *Distinguishing Between Web Data Mining And Information Access*. Presentation For The Panel on Web Data Mining, kdd 97, August 16, 1997, Newport Beach, ca. Retrieved March 18, 2013, from http://www.sims.berkeley.edu/

Hearst, M. A. (1999, June). Untangling text data mining. In Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics (pp. 3-10). Association for Computational Linguistics.

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics, 52*(2), 291-314.

Hotho, A., Nürnberger, A., & Paaß, G. (2005, May). A Brief Survey of Text Mining. *In Ldv Forum, 20*(1), 19-62. Retrieved November 3, 2013, from http://www.kde.cs.uni-kassel.de/hotho/pub/2005/hotho۰۵TextMining.pdf

Hudson, J. (1996). Trends in multi-authored papers in economics. *Journal of Economics Perspectives, 10*, 153–158.

Janssens, F., Leta, J., Glänzel, W., & De Moor, B. (2006). Towards mapping library and information science. *Information Processing & Management, 42*(6), 1614-1642.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration?. *Research Policy, 26*(1), 1-18.

Kontardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. IEEE Press.

Kostoff, R. N., & Demarco, R. A. (2001). Information Extraction From Scientific Literature With Text Mining. *Journal of Analytical Chemistery, 73*(4), 370-378.

Li, L. L., Ding, G. H., Feng, N., Wang, M. H., & Ho, Y. S. (2009). Global stem cell research trend: Bibliometric analysis as a tool for mapping of trends from 1991 to 2006. *Scientometrics, 80*(1), 39–58.

Lowe, M. (2003). Reference analysis of the American historical review. *Collection building, 22* (1). Retrived November 18, 2013, from http://www.

emeraldinsight.com/Insight/viewPDF.jsp?Filename=html/Outpt/Published/EmeraldFULLTextArticle/pdf/1710220103.pdf

Lundberg, J., Tomson, G., Lundkvist, I., Skor, J., & Brommels, M. (2006). Collaboration uncovered: exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. *Scientometrics, 69*(3), 575–589.

Marshakova-Shaikevich, I. (2005). Bibliometric maps of field of science. *Information Processing & Management, 41*(6), 1534-1547.

Moed, H. F., Glanzel, W., & Schmoch, U. (2004) (eds.). *Handbook of quantitative science and technology research.* The use of publication and patent statistics in studies of S&T systems. Dordrecht (the Netherlands): Kluwer Academic Publishers, 800 pp.

Noyons, E. C. (1999). *Bibliometric Mapping as a Science Policy and Research Management Tool*. Leiden: DSWO Press.

Noyons, E. C. (2001). Bibliometric mapping of science in a policy context. *Scientometrics, 50*(1), 83-98.

Polanco, X., & Zartl, A. (1999). *Information visualization*. EICSTES Project-IST. Deliverable 1.4 State of the art WP9: visualization. Retrieved May 14, 2013, from www.eicstes.org/EICSTES_PDF/Deliverables/Information%20Visualization. Pdf

Pons-Porrata, A., Berlanga-Llavori, R., & Ruiz-Shulcloper, J. (2007). Topic discovery based on text mining techniques. *Information processing & management, 43*(3), 752-768.

Rübel, O., Ahern, S., Bethel, E., Biggin, M. D., Childs, H., Cormier-Michel, E., ... & Wu, K. (2010). Coupling visualization and data analysis for knowledge discovery from multi-dimensional scientific data. *Procedia computer science,1*(1), 1757-1764.

Sharma, N. (2005). *Discovering knowledge with text mining*. M. S. Thesis, Texas A&M University.

Shiffrin, R. M., & Borner, K. (2004). Introdoction. In: Mapping knowledge domains. *PNAS, 101*, Suppl. 1, 5183-5185. Retrieved September 30, 2013, from www.pnas.org/cgi/doi/10.1073/pnas.0307852100

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American society for Information Science, 50*(9), 799-813.

Small, H. (2000). Charting pathways through science: Exploring Garfield's

vision of a unified index to science. *The web of knowledge: A Festschrift in honor of Eugene Garfield*, 449-473.

Small, H.)1973(. Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science and Technology, 24*(4), 265–269.

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: identifying Bibliography 215 and graphing specialties. *Science Studies, 4*(1), 17-40.

Swanson, D. R., & Smalheiser, N. R. (1999). Implicit text linkages between Medline records: Using Arrowsmith as an aid to scientific discovery. *Library trends, 48*(1), 48-59.

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.

Taylor, P. (2006). *From patient data medical knowledge: the principles & practice of health informatics*. Massachusetts: Black Well.

Thomas, J., McNaught, J., & Ananiadou, S. (2011). Applications of text mining within systematic reviews. *Research Synthesis Methods, 2*(1), 1-14.

Turban, E., Mclean, E., & Wetherbe, J. (1999). *Information technology for management: making connections for strategic advantage*. New York: Willey & sons.

Van Eck, N. J., & Waltman, L. (2007). VOS: A new method for visualizing similarities between objects. In H.-J. Lenz & R. Decker (Eds.), *Advances in data analysis*:*Proceedings of the 30th annual conference of the german classification society*, (pp. 299-306). Springer.

Walter, M., Stuart, L., & Borisyuk, R. (2004). The representation of neural data using visualization. *Information Visualization, 3*(4), 245-256.

Wang, Y., Wu, Y. S., Pan, Y. T., Ma, Z., & Rousseau, R. (2005). Scientific collaboration in China as reflected in co-authorship. *Scientometrics, 62*(2), 183–198.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for information Science, 32*(3), 163-171.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-

citation analysis of information science, 1972-1995. *Journal of the American society for in formation science, 49*(4), 327-355.

Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *Journal of High Technology Management Research, 15*(1), 37–50.