

کاربردهای کتاب‌سنجی و تحلیل پیوند

در محیط وب

فریبرز درودی*

دکترای کتابداری و اطلاع‌رسانی
سازمان اسناد و کتابخانه ملی ایران

اطلاعات
علمی و فناوری

دریافت: ۱۳۸۷/۰۷/۱۶ | پذیرش: ۱۳۸۸/۰۱/۱۹ | مقاله برای اصلاح به مدت یک ماه و ۱۲ روز نزد پدیدآور بوده است.

چکیده: هدف از انجام این پژوهش، معرفی و تحلیل کاربردهای کتاب‌سنجی در محیط وب و نیز تبیین وضعیت تحلیل پیوند به منظور بیان کاربردهای آن در ارتباط با منابع اطلاعاتی موجود در وب است. روش این پژوهش مطالعه، بررسی و تحلیل منابع اطلاعاتی در حوزه کاربرد کتاب‌سنجی و تحلیل پیوند در وب، به منظور تبیین وضعیت آن است. نتایج پژوهش نشان می‌دهد که کتاب‌سنجی می‌تواند کاربرد مطلوبی در حوزه منابع دیجیتال موجود در اینترنت داشته باشد. همچنین تحلیل پیوند، فرایندی است که با تکیه بر آن می‌توان به تحلیل آماری ارتباط میان فرآیندها پرداخت و در نتیجه به صحت و سقم، و اثربخشی استنادهای موجود در مدارک دیجیتال پی برد. در واقع تحلیل پیوند، بخشی از الگوریتم رتبه‌بندی اطلاعات در محیط وب به‌شمار می‌آید. تعداد، ربط و کیفیت پیوندهای معین به یک وب‌سایت از عوامل بسیار مهم رتبه‌بندی/ تثبیت موقعیت در محیط وب می‌باشد. ابزارهای کاربردی در این زمینه موضوعی شامل راهبرد رتبه- صفحه، الگوریتم‌های تحلیل پیوند، نمایه‌سازی معنایی پنهان، مدل کلاسیک درون‌داد- برون‌داد می‌شود. در این پژوهش وضعیت تحلیل پیوند در وب جهانی و وب کوچک تحلیل شده و شیوه بهره‌گیری از نمودارهای وب برای درک بهتر از فرایند تحلیل پیوند، توضیح داده شده است.

کلیدواژه‌ها: کتاب‌سنجی؛ تحلیل پیوند؛ وب‌سنجی؛ رتبه‌بندی؛ وب جهان‌گستر

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شبا (چاپی) ۵۲۰۶-۱۷۳۵
شبا (الکترونیکی) ۵۵۸۳-۲۰۰۸
نمایه در SCOPUS و LISA
<http://jlist.irandoc.ac.ir>
دوره ۲۵ | شماره ۳ | صص ۵۴۹-۵۶۷
بهار ۱۳۸۹

نوع مقاله: علمی مروری

*fardoroudi@gmail.com

۱. مقدمه

با پیشرفت‌های نوین فناوری در عرصه ارتباطات مجازی و محیط شبکه، دسترسی به اطلاعات، بیش از پیش با سهولت افزون‌تر مهیا شده. این در حالی است که با افزایش اطلاعات و عدم سازماندهی مدون و یکپارچه در محیط وب، با تراکم حجم اطلاعات در آن و امکان بازیابی مقادیر بسیار زیاد اطلاعات مرتبط با حوزه‌های مطالعاتی متعدد روبرو هستیم. این پدیده از یک سو امکان بازیابی اطلاعات مستند و معتبر را دشوار ساخته، و از سوی دیگر فعالیتی مضاعف برای ارزیابی و تحلیل اعتبار اطلاعات بازیابی شده را می‌طلبد. فعالیت‌های کتاب‌سنجی در محیط شبکه، و بخصوص در عرصه وب، یکی از فعالیت‌های ضروری و لازم برای دستیابی به اطلاعات معتبر، ارزشیابی، و سنجش اعتبار این اطلاعات به شمار می‌آید. کتابداران برای دستیابی به این هدف، نیازمند آشنایی با فرایندهای کتاب‌سنجی در محیط شبکه می‌باشند. بر این اساس، ضروری است که با بهره‌گیری از معیارهای مدون، به سنجش ارتباط میان عناصر اطلاعاتی در وب پرداخت و با این شیوه از ارزش و اعتبار منابع اطلاعاتی و هسته اصلی این منابع موضوعی در شبکه آگاهی یافت. بر همین اساس، مطالعه و بررسی در حیطه موضوعی تحلیل پیوند^۱ در محیط وب، یکی از عرصه‌های ارزیابی این مقوله مهم می‌باشد. شناخت فرایندهای کتاب‌سنجی در رابطه با صفحات الکترونیکی و مطالعه در باب ارتباطی که میان آن‌ها برقرار می‌شود، می‌تواند به نتایج سودمندی در این زمینه منجر شود.

۲. مفهوم کتاب‌سنجی

«لاندبرگ» بیان می‌کند که کتاب‌سنجی (bibliometrics) کاربرد روش‌های ریاضی و آماری برای تبیین وضعیت انتشارات است (Lundberg 2006). این اصطلاح از ریشه دو واژه کهن "biblos" به معنای کتاب و "metron" یعنی اندازه‌گیری، اخذ شده است. کتاب‌سنجی اغلب برای ارزیابی پژوهش‌های علمی از طریق مطالعات کمی در انتشارات پژوهشی مورد استفاده قرار می‌گیرد. «جاکوبس» اظهار می‌دارد که مطالعات مربوط به ویژگی‌های انتشارات که بانام کتاب‌سنجی یا مطالعات کمی شناخته شده، معیار اثربخشی برای سنجش فرآورده‌های علمی، گرایش‌ها، اهمیت پژوهش در شاخه‌های متعدد علمی، و سلیقه‌های

¹ link analysis

پژوهشگران برای تولید انتشارات می‌باشد (Jacobs 2001). «گلنزل» نیز اعلام می‌دارد که اصطلاح کتاب‌سنجی و علم‌سنجی^۱ تقریباً به صورت همزمان توسط «پریچارد»^۲، «نالیموف»^۳، و «مالچنکو»^۴ در ۱۹۶۹ مطرح شد. در حالی که «پریچارد» توضیح داد که اصطلاح کتاب‌سنجی عبارت است از کاربرد سنجش روش‌های ریاضی و آماری برای کتاب‌ها و دیگر رسانه‌های ارتباطی، «نالیموف» و «مالچنکو» علم‌سنجی را کاربرد روش‌های کمی که با تحلیل علم (به عنوان پردازش اطلاعات) ارتباط دارد، دانستند (Glanzel 2003).

«آرچمبولت» و «گانگ» بیان می‌کنند که کتاب‌سنجی و علم‌سنجی، مجموعه‌ای از روش‌ها برای سنجش تولیدات، فرآورده‌ها و اشاعه دانش علمی هستند (Archambault and Gange 2004). اما تعریفی که دایره‌المعارف ویکیپدیا ارائه می‌دهد چنین است: کتاب‌سنجی مطالعه، یا فن اندازه‌گیری متون و اطلاعات است (Wikipedia 2008).

با بهره‌گیری از مفهومی که پژوهشگران در حوزه کتاب‌سنجی از این زمینه موضوعی ارائه داده‌اند، می‌توان کتاب‌سنجی را شیوه‌های بررسی و سنجش تولیدات علمی، عمدتاً با تکیه بر روش‌های کمی، و تحلیل میزان ارتباط میان عناصر اطلاعاتی منتشر شده دانست. با بهره‌گیری از روش‌های کتاب‌سنجی میزان مستندات، اعتبار، و اثربخشی متون علمی مشخص می‌گردد، و میزان اثربخشی منابع اطلاعاتی تعیین می‌شود.

۳. اجزا، عناصر، و تأثیر کتاب‌سنجی

«گلنزل» بیان می‌کند که امروزه کتاب‌سنجی یکی از حوزه‌های پژوهشی بین‌رشته‌ای واقعاً بی‌نظیر برای توسعه اغلب رشته‌های علمی است. روش‌شناسی کتاب‌سنجی شامل عواملی از ریاضیات، علوم اجتماعی، علوم طبیعی، مهندسی، و حتی علوم زیستی است (Glanzel 2003). «آرچمبولت» و «گانگ» نیز اظهار می‌دارند که کتاب‌سنجی بر پایه دو پیش‌فرض بنا نهاده شده است: اول، هدف پژوهشگران پیشرفت و توسعه دانش است، و این بدان معنا است که اشاعه نتایج مطالعات و پژوهش‌های آنان از طریق انواع رسانه‌های

¹ scientometrics

² Pritchard

³ Nalimov

⁴ Mulchenko

ارتباطی (شامل نگارش‌هایی که در مرکز و کانون توجه جریان علمی واقع شده‌اند) انجام می‌پذیرد. دوم، دانشمندان ناچار به انتشار نوشته‌های خود هستند تا بتوانند به اعتبار و پیشرفت علمی دست یابند (Archambault and Gange 2004).

از سوی دیگر «گلنزل» نیز مطرح می‌کند که سه مؤلفه برای مفهوم کتاب‌سنجی در دوران معاصر وجود دارد: اول، کتاب‌سنجی برای کتاب‌سنج‌ها (روش‌شناسی). این مورد، دامنه‌ای از پژوهش‌های پایه کتاب‌سنجی است. پژوهش‌های روش‌شناسانه به طور کلی در این حوزه هدایت می‌شوند؛ دوم، کتاب‌سنجی برای رشته‌های علمی (اطلاعات علمی)، که این حیطه ممکن است به عنوان توسعه و گسترش اطلاعات علمی توسط روش‌های قیاسی مطرح شود. در اینجا ما همچنین ارتباط تنگاتنگی با پژوهش‌های کمی در عرصه بازیابی اطلاعات مشاهده می‌کنیم؛ سوم، کتاب‌سنجی برای خط‌مشی و مدیریت علمی (سیاست علمی)، که دامنه‌ای از سنجش و ارزشیابی پژوهش می‌باشد و در حال حاضر مهم‌ترین موضوع در این حوزه است (Glanzel 2003).

همچنین «لاندرگ» با اشاره به ارتباطات میان دو عنصر مطرح در کتاب‌سنجی بیان می‌کند که تعدادی از استندهایی که به یک مقاله در مجله علمی می‌شود، می‌تواند به عنوان بازتاب تأثیر مقاله بر مجامع علمی مورد بررسی قرار گیرد (Lundberg 2006). «برتین» و دیگران نیز بیان می‌کنند که کتاب‌سنجی، ارزیابی کمی مطالب نوشتاری است. مطالعات بسیاری برای توسعه این علم و کشف مقیاس‌های پذیرش برای برآورد تولیدات علمی پژوهشگران در یک کشور یا یک مؤسسه مورد استفاده قرار می‌گیرند (Bertin et al 2006).

بر مبنای ارتباطات پدیدآمده میان اجزای مهم کتاب‌سنجی که در بالا مورد اشاره واقع شد، پیوندهای استنادی شکل می‌گیرد. میزان تأثیر عناصر استنادی در متون علمی، با تحلیل راهبردهای کتاب‌سنجی به انجام می‌رسد. با استفاده از روش‌های تحلیل استنادی خاص کتاب‌سنجی می‌توان تأثیر متون علمی بر یکدیگر را مورد محاسبه قرار داد، و میزان بهره‌گیری نویسندگان از آثار یکدیگر را معین ساخت.

۴. کاربرد کتاب‌سنجی در محیط وب

از آنجا که حیطه وسیعی از اطلاعات در محیط وب وجود دارد که به صورت بالقوه در تمامی جنبه‌های زندگی اجتماعی ما مؤثر است، به دست آوردن معیار ربط^۱ صفحات الکترونیکی برای تأمین نیاز اطلاعاتی کاربر، هنوز یک وظیفه چالش‌برانگیز و مهم محسوب می‌شود (Ding 2003). این امر تا حد زیادی به مقوله «اعتبار اطلاعات دیجیتال در محیط شبکه» ارتباط دارد. با گسترش فضای وب و ورود اطلاعات متنوع و متکثر در این فضای مجازی، به علاوه فقدان نظارت بر درستی و نیز مستند بودن اطلاعات، انجام مطالعات کتاب‌سنجی، به همراه پژوهش‌های مرتبط با مقوله تحلیل پیوندهای برقرار شده میان صفحات الکترونیکی، از اهمیت بیش تری برخوردار است. در این میان، فرایوند^۲ نقش اثرگذاری برای سنجش این مقوله دارد.

فرایوند در بین وب‌سایت‌ها، استادهای مقالات به یکدیگر را مورد مقایسه قرار می‌دهد و این فرایند سبب ایجاد تعدادی از مطالعات اکتشافی برای اثبات آن شده است؛ چراکه روش‌های کتاب‌سنجی مرسوم می‌تواند به صورت عملی در تمام، یا بخشی از محیط وب این کار را انجام دهد (Archambault and Gange 2004). این شیوه‌های کاربردی در محیط شبکه نیاز به بهره‌گیری از برخی روش‌های خاص فضای مجازی در محیط وب دارد. بر این اساس سنجش پیوندهای برقرار شده از طریق فرایوند، و کنترل استنادی آن به شکل خاص در ارتباط با مقالات الکترونیکی که عمدتاً توسط مجلات علمی مورد استفاده قرار می‌گیرد، یکی از راهکارهای مؤثر در این عرصه است.

مطالعات کتاب‌سنجی می‌توانند به عنوان پایه‌ای برای اصطلاحات نمایه‌ای، واژه‌هایی در عناوین مدارک، پدیدآورندگان منابع، یا توزیع زمانی و مکانی به کار روند. به طور نمونه، «نمایه استنادی علوم»^۳ و «نمایه استنادی علوم اجتماعی»^۴ که نمونه‌های شناخته شده و معروفی در این زمینه هستند، از این رویکردها بهره می‌گیرند (Stenberg and Ungern 1995). راهبردهای استفاده شده در چنین منابع اطلاعاتی، کاربردهای وسیعی در زمینه مطالعات استنادی و تعیین اعتبار اطلاعات بر عهده دارند.

¹ Relevance

² Hyperlink

³ Science Citation Index

⁴ Social Science Citation Index

کتاب‌سنجی به شکل کاربردی برای جویاشدن پرسش‌های جدید، در شیوه‌های نوین نمایان شده است. تحلیل هم‌استنادی^۱ برای شناخت روابط میان انتشارات چاپی به صورت عملی، در چارچوب فضای ذهنی وب طرح شده است (Larson 1996, Qouted in Borgman and Jonathan 2002) و قابلیت بهره‌گیری اثربخش در محیط دیجیتال را دارا می‌باشد.

باید بیان کرد که کتاب‌سنجی به‌طور کامل قابلیت اجرا در محیط وب، یعنی بزرگ‌ترین شبکه مدارک در طول تاریخ را دارد. ما می‌توانیم از فنون، قوانین و قواعد کتاب‌سنجی در محیط پایه (اطلاعاتی درباره اطلاعات سازماندهی شده در نظام‌های اطلاعاتی) بهره‌گیریم (Turnbull 1997) و از آن به‌عنوان ابزاری اثرگذار در شبکه یاری جوییم.

۵. فرایند تحلیل پیوند در وب

ایده اصلی جستجوی وب، یافتن مجموعه صفحات رتبه‌بندی شده بر مبنای معیار ربط، بر اساس اندیشه یا مفهومی است که استفاده‌کننده در ذهن دارد. این مفاهیم بر اساس کلیدواژه‌هایی که باید در هر مدرک بر مبنای ربط مفروض ظاهر شود، مشخص شده است (Kohlschütter, Chirita and Nejd1 2001) و ربط میان واژگان، با مفهوم پیوند متبلور می‌شود. اصطلاح‌های پیوند و پیوند دادن، به شیوه علمی، حرفه‌ای، و نیز همانند نوشته‌های مردم‌پسند، با فرامتن به صورت کلی، و وب جهان‌گستر به صورت خاص، وابسته هستند. از دیدگاه تاریخی، متخصصان کتاب‌سنجی علاقه‌مند هستند که به‌طور صریح شرح وابستگی میان اسناد را با مباحث مربوط به استناد و استناد دادن بیان کنند. این اصطلاحات، و حوزه مطالعه آن با عنوان تحلیل استنادی شناخته شده است (Borgman and Jonathan 2002). تحلیل استنادی به بررسی رابطه معنایی که میان متن و منابع استنادی آن وجود دارد، می‌پردازد و تلاش می‌کند قوانین سازنده روابط استنادی را معین و مشخص سازد. این فرایند در حوزه وب و اطلاعات دیجیتال نیز از اهمیت خاص برخوردار است. می‌توان با فنون تحلیل استنادی به بررسی دقیق اثرپذیری اطلاعات موجود در محیط شبکه پرداخت. فرآیندهای موجود در صفحات وب یکی از عناصری هستند که در این میان،

¹ Co-citation

متخصصان موضوعی از آن یاری می‌جویند و با طرح پرس‌وجوهای معین، به بررسی میزان ربط آن‌ها مبادرت می‌ورزند. طرح پرس‌وجو در محیط وب بر اساس ارتباط مفهومی میان کلیدواژه‌های انتخاب‌شده می‌باشد. برای این کار باید از برخی کلیدواژه‌های مشابه در پرس‌وجوهای نمونه بهره گرفت.

به‌طور خلاصه ارائه پرس‌وجوی مشابه شامل موارد زیر می‌شود: تعمیم‌دادن از مدرک خاص با استخراج ویژگی‌های مرتبط با آن؛ تحلیل گروهی از مدارک برای انطباق بر آن خصوصیات؛ رتبه‌بندی مدارک انتخاب‌شده با در نظر گرفتن ویژگی‌هایی که از مجموعه مدارک، حاصل آمده است؛ ارائه مقداری از مدارک رتبه‌بندی‌شده به‌عنوان بیش‌ترین مشابهت (Schiffman 2006). با این تدبیر، برخی نتایج حاصل از ارتباط میان اجزای اطلاعاتی در محیط وب مشخص می‌شوند و می‌توان بر مبنای آن به بررسی ربط مدارک و فرایوندهای ایجادشده مبادرت ورزید. در بیش‌تر مواقع، این فعالیت به نتایجی منجر می‌شود که با محاسبه آماری ارتباطات میان فرایوندها، می‌توان به صحت و سقم، و اثربخشی استادهای موجود در مدارک دیجیتالی پی برد.

۶. ابزارهای سنجش و کاربرد فرایند تحلیل پیوند در محیط وب

علاقه حرفه‌ای در فعالیت انتشاراتی که دانشمندان انجام می‌دهند - در واقع از سوی نویسندگان عمومی - می‌تواند کم‌تر یا بیش‌تر از نمونه دقیق، مأخوذ از نیم‌رخ باشد که به‌صورت خودکار از اصطلاح‌های وابسته با نویسندگان‌شان در پایگاه اطلاعاتی، به‌وجود می‌آید. فناوری موجود می‌تواند فرایند حرفه‌ای آن را برای هر کسی در جهت ساختن چنین نمایی، ساده سازد. به خدمت در آوردن نرم‌افزارهای مناسب، انواع جستجوکننده‌ها در مجموعه نوشته‌ها توسط نویسنده، و سپس فراخوانی اصطلاح‌ها (عبارات اسمی از یک نوع خاص) که به‌صورت همزمان با نام‌های ورودی در پیشینه‌های کتابشناختی، از بالا به پایین توسط بسامد همزمان رویدادها رتبه‌بندی می‌شود (White 2001)، از راهبردهای مؤثر در فرایند تحلیل پیوند به‌شمار می‌آید. این بررسی‌ها نتیجه مطلوبی از بررسی وضعیت تولیدات علمی به تصویر می‌کشد.

¹ query

۶-۱. الگوی فرامتن مکانی

یکی از ابزارهای نوین تحلیل پیوند، الگوی فرامتن مکانی است که منجر به گزینش بهینه موضوع می شود. بر اساس نظر «کلینبرگ»^۱، این فن جدید به بررسی و تحلیل نمودار معنایی می پردازد، که از متون تثبیت شده ای که با ساختار فرایوند وب کار می کنند تشکیل شده است. متون تثبیت شده، برای جستجو در محیط وب سودمند است، زیرا به مثابه یک عنوان مورد توافق جمعی عمل می کند. ساختار مجاورت یک نمودار معنایی نمی تواند به عنوان یک ماتریس، بدون از دست دادن اطلاعات حاشیه ای، مدل سازی شود (Kolda, Bader, and Kenny 2005) و همین ویژگی، کاربرد آن را تا حدی با دشواری همراه ساخته است. برای بهره گیری از روش مذکور می توان با تحلیل های ساختاریافته در محیط وب، بر اساس بُردارهای تنظیم شده در فضای مجازی، و با استفاده از مقیاس های اندازه گیری استاندارد، به تبیین وضعیت آن پرداخت. بر همین اساس، بهره گیری از الگوریتم های خاص این راهبرد مورد توجه قرار می گیرد.

۶-۲. الگوریتم های تحلیل پیوند (راهبرد رتبه- صفحه)

الگوریتم های تحلیل پیوند^۲ نظیر راهبرد «رتبه- صفحه»^۳ و «فرایوند حاصل از جستجوی موضوعی»^۴، در محاسبات بُردار ویژه^۵ به منظور تعیین صفحات معتبر، به عنوان پایه ای در ساختار فرایوند مورد استفاده قرار می گیرند (Xue et al 2003). راهبرد رتبه- صفحه، فرمولی است که توسط گوگل^۶ توسعه داده شد تا در تعیین رتبه بندی پیوندهای مشخص در صفحات وب مورد بررسی قرار گیرد. در واقع روش انحصاری گوگل برای اندازه گیری اهمیت صفحه وب، به کارگیری معیار ارزش گذاری بر مبنای امتیاز صفر تا ده می باشد. امتیاز ده بالاترین رتبه ای است که ممکن است به صفحه ای تعلق گیرد. الگوریتم این راهبرد، کمیت و کیفیت پیوندهایی که به صفحات وب تعلق گرفته را مورد تحلیل قرار می دهد. با بهره گیری از این روش می توان میزان ارتباط صفحه الکترونیکی موجود در

¹ Kleinberg² link analysis algorithm³ PageRank⁴ Hyperlink Induced Topic Search(HITS)⁵ Eigenvector calculations⁶ www.google.com

محیط وب با پیوند ایجاد شده را مشخص ساخت. از این راهبرد سودمند می‌توان به خوبی در ارتباط با تعیین اعتبار پیوند که از منظری دیگر به اطلاعات آن صفحه مربوط می‌شود، بهره گرفت.

از سوی دیگر، با تمرکز بر رویکرد خاص استخراج داده، مستقیماً از اطلاعات گزارمانی^۱ برای پشتیبانی از جستجوی چهریزه‌ای در نمودارهای مقیاس وسیع وب، که اصطلاحاً رتبه - صفحه اختصاصی^۲ خوانده می‌شود، بهره می‌گیریم. نسخه ویژه شناخته شده رتبه - صفحه در صفحات وب، محبوبیت زیادی در فرایند سنجش کسب کرده است (Kohlschütter, Chirita and Nejd1 2006) و معیاری کاربردی برای ارزیابی فرایوند به‌شمار می‌آید. علاوه بر آن، استفاده از مدل در زمینه تحلیل پیوند، برای نمایش جریان حرکت و سنجش فرایندهای ارتباطی آن، از تأثیر بسزایی برخوردار است.

۶-۳. نمایه‌سازی معنایی پنهان

گوگل برای دستیابی به مرتبط‌ترین نتایج در جستجوهای موضوعی، از روش نمایه‌سازی معنایی پنهان^۳ بهره گرفته است. نمایه‌سازی معنایی پنهان، سبب بهینه‌سازی عملکرد موتورهای جستجو شده است. این فرایند به این معنا است که موتور جستجو تلاش می‌کند در هنگامی که به نمایه‌سازی صفحات وب می‌پردازد، اصطلاح‌های معین و مشخصی که به آن ارائه شده را با مفاهیم و معنایی که با آن‌ها سازگاری دارند، مرتبط سازد. به‌عنوان مثال به جای آن که به ایجاد ربط میان نام «پاریس هیلتون»^۴ با نام شهر یا اسم هتل بپردازد، آن را با مفهوم زن مرتبط بداند. گوگل در حال حاضر در بسیاری از خدمات مرتبط با آگهی‌های خود از این شیوه بهره گرفته است. همچنین گوگل هم‌اکنون در الگوریتم جستجوی خود کوشش کرده است این ایده در نتایج جستجو ظاهر شود.

الگوریتم نمایه‌سازی معنایی پنهان، عبارت‌سازی در صفحات و تعیین موضوع آن صفحات را انجام می‌دهد. ولی فقط پس از آن که گوگل از این الگوریتم برای شیوه جستجوی موتور جستجوی خود بهره گرفت، توسط دیگر موتورهای جستجو مورد

¹ annotated information

² Personalized PageRank

³ Latent Semantic Indexing (LSI)

⁴ Paris Hilton

بهره‌برداری قرار گرفت. این روش، تحلیل و بررسی واژه‌های به کار رفته در زبان طبیعی را شامل می‌شود. مترادف‌ها و واژگان مرتبط و وابسته، در هنگام بحث از موضوع‌های کلی در یک صفحه وب مورد استفاده قرار می‌گیرند. این فرایند با کامل کردن، به جای تعویض واژگان به تحلیل کلیدواژه‌ها می‌پردازد (Nisbet 2006). با ایجاد چنین زمینه مناسبی در برقراری ربط مناسب کلیدواژه جستجو با نتایج بازیابی شده، می‌توان انتظار داشت که پیوندهای ایجاد شده، قابلیت تطبیق بیش تری با مفاهیم مورد نظر داشته باشند و از معیار رتبه - صفحه به صورت بهینه در این عرصه بهره گرفت. معیار رتبه - صفحه، قابلیت ارتقای بیش تر در جستجوهای محیط وب دارد.

۶-۴. مدل کلاسیک درونداد-برونداد

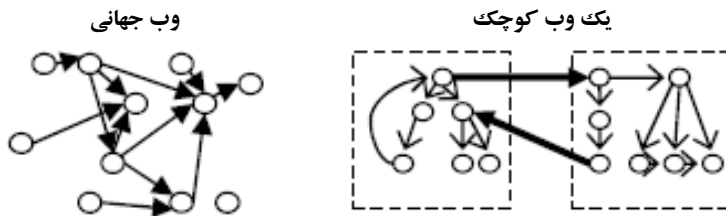
مدل کلاسیک درونداد-برونداد^۱ برای توصیف فرایند پژوهش علمی مورد استفاده قرار می‌گیرد. انتشارات معمولاً در قالب مقالات داوری شده و تک‌نگاشت‌های علمی (به‌مثابه اظهارات نهایی نتایج طرح پژوهش) توجه می‌کنند. این محصولات اطلاعاتی می‌توانند برای تعیین اندازه و درک ماهیت پژوهش، ارزش‌گذاری و تحلیل شوند. سنجش میان ورودی و خروجی اطلاعات علمی، به‌عنوان روشی مؤثر و مورد قبول، در مجامع علمی پذیرفته شده است.

تحلیل کتاب‌سنجی در فعالیت‌های علمی، پایه‌ای بر این فرضیه است که درک پژوهش و برقراری ارتباط، نتیجه‌ای مشترک خواهد داشت. دستیابی به پیشرفت علمی توسط پژوهشگران، همراه با مطالعه مباحث پژوهشی خاص، و هدایت توسط آثار پیشین همکاران پژوهشگر حاصل می‌آید. مقیاس اندازه‌گیری کتاب‌سنجی در سطوح بالای تراکم از قدرت زیادی برخوردار است و برای تحلیل الگوها در مجموعه‌های بزرگ (گروه‌های دانشگاهی یا تیم‌های پژوهشی بزرگ) موقعیت بسیار مناسبی دارد، و برای ارزشیابی افراد یا گروه‌های پژوهشی کوچک مناسب نیست. بر این اساس، اعتبار مقیاس کتاب‌سنجی هنگامی که برای مجموعه‌های داده‌ای کوچک به کار رود، به نظر مشکوک می‌آید و بررسی نقادانه و داوری مجدد در این سطح الزامی است و باید به سطح و محدوده استفاده از این ابزار اثربخش، توجه جدی داشت (Russell and Rousseau 2002).

¹ classic input-output model

۷. تحلیل پیوند در وب جهانی و وب کوچک

در بررسی تداوم رشد ابعاد وب، باید روش‌های تحلیل پیوند برای توسعه آن مورد ملاحظه قرار گیرد (Kolda and Bader 2006) و به صورت گسترده، از آن در فضای شبکه استفاده مناسب به عمل آید. فرایند جستجو در فضای وب، یا در حیطه گسترده‌ای از وب می‌باشد که آن را به نام وب جهانی^۱ می‌شناسیم، یا در عرصه محدودی از آن انجام می‌شود که آن را وب کوچک^۲ نام گذارده‌اند. «زو» و دیگران بیان می‌کنند که منظور از وب کوچک، آن بخشی از اینترنت است که در فضای محصور در یک سازمان یا مؤسسه‌ای خاص، به ارائه خدمات می‌پردازد- نظیر اینترانت^۳ که از لحاظ حوزه جغرافیایی و محدودیت محیط شبکه، کوچک‌تر از وب جهانی است (Xue et al 2003).



شکل ۱. ارتباطات میان گره‌ها و پیوندهای موجود در محیط وب (Xue et al 2003)

سنجش معیارهای تحلیل پیوند در محیط وب، چه وب جهانی و چه وب کوچک، از اهمیت زیادی برخوردار است. بهره‌گیری از الگوریتم‌های خاص محیط وب، برای تحلیل پیوندهای ارائه شده، بسیار سودمند است. میزان ارتباط میان پیوندهای پدیدآمده در فضای وب، بر اساس نوع وب- جهانی یا کوچک- با تغییراتی همراه است. ارتباطات میان پیوندها با ترسیم نمودارهای مرتبط با گره‌ها و ربط خاص آن‌ها قابل نمایش است. این گره‌ها از طریق پیوندهای مرتبط و برقراری رابطه اثرپذیری، به ارائه اطلاعات مفیدی در این عرصه می‌انجامد.

¹ Global Web

² Small Web

³ Intranet

فرضیه پایه‌ای وجود دارد مبنی بر این که الگوریتم تحلیل پیوند، در توصیف وب جهانی دارای یک نمودار یکپارچه استنادی است (طرح سمت چپ از شکل ۱)، و هر فرایوند نمایش داده شده، یک استناد یا ارتباط توصیه شده به همراه دارد. باید گفت که فرضیه پیشنهاد شده عبارت است از آن که یک فرایوند در صفحه X ، به نشانه توصیه‌ای که نویسنده صفحه X درباره صفحه Y انجام داده، دلالت بر صفحه Y دارد. بر اساس فرضیه پیشنهادی در عرصه وب جهانی، به دلیل این که فرایوندها بر اساس قضاوت نویسندگان، رمزگذاری و تحدید می‌شوند، معمولاً مطابق با واقعیت است. البته برخی از فرایوندها برای مقاصد توصیه شده ایجاد نشده‌اند، اما تأثیر آن‌ها می‌تواند با اعمال فرایند فیلتر یا مراتب نادیده انگاشتن آن، کاهش یابد.

این فرضیه پیشنهادی، در مورد محیط وب کوچک قابل سنجش نیست (Xue et al 2003)؛ زیرا وب کوچک از لحاظ ساختار جغرافیایی و فضای جستجو، متفاوت با وب جهانی است. این محیط بر اساس نوع اطلاعات و همگنی میان آن‌ها، با خصوصیات متفاوتی ظاهر شده است. تعداد پدیدآورندگان، محدود است و عمدتاً با توجه به تعداد رابطه برقرار شده میان آثار آنان، از بسامد و تکرار بیش‌تری نسبت به پیوندهای ایجاد شده، برخوردارند. در مواردی نیز استنادها توسط نویسندگان به آثار خودشان انجام می‌پذیرد، که سبب تضعیف اعتبار فرایوند در محیط وب کوچک است. با این حال، امکان خطاپذیری در محیط وب کوچک به مراتب بیش‌تر می‌باشد.

همان‌طور که در قسمت راست شکل ۱ مشاهده می‌شود، بیش‌تر فرایوندها در محیط وب کوچک، نسبت به وب جهانی از نظم و قاعده‌مندی بیش‌تری برخوردارند. بیش‌تر پیوندها در حرکت از گره مادر به سوی گره‌های فرزند^۲، بین گره‌های هم‌سان^۳، یا در حال ترک ریشه هستند. دلیل آن به‌طور عمده آن است که فرایوندها در وب کوچک، توسط بخش کم‌تری از نویسندگان تولید شده، و هدف فرایوندها معمولاً سازماندهی محتوا، در یک ساختار خطی یا سلسله‌مراتبی است. بنابراین سنجش اعتبار، بازتاب معتبر بودن صفحه نمی‌باشد؛ بلکه باید اظهار داشت که ایجاد الگوریتم تحلیل پیوند موجود، برای محیط وب

¹ parent node

² children nodes

³ sibling nodes

کوچک مناسب نیست. فرایوند [در محیط وب] می تواند به دو گروه پیوندهای هدایتی^۱ و پیوندهای پیشنهادی^۲ تقسیم شود. مورد اخیر برای فرایند تحلیل پیوند به منظور انجام جستجوهای زیاد، مناسب تر است (Xue et al 2003). البته باید توجه داشت که در وب کوچک، فرایوند را نمی توان به پیوندهای هدایتی و پیشنهادی تقسیم کرد، که این یکی از محدودیت های محیط وب کوچک است.

۸. بهره گیری از نمودارهای وب برای درک بهتر از تحلیل پیوند

برای نمایش تصویری ارتباطات میان گره ها و فرایوندهای ارائه شده در محیط وب، می توان از نمودارهای وب بهره گرفت که کاربرد مؤثری در میزان درک بهتر از روابط ایجاد شده دارند. این نمودارها در حالت های متفاوت، ارتباطات میان فرایوندها را به نمایش می گذارند. به کمک چنین نمودارهایی می توان به تحلیل مناسب پیوندهای برقرار شده پرداخت و روابط مدون میان آنها را معین ساخت. در این بین با بهره گیری از دو منحنی توزیع در یک صفحه واحد، وضعیت پیوندها بهتر قابل نمایش است. نمودار هیستوگرام وب که با تعیین توزیع درجه ورودی و خروجی به نمایش در می آید، یکی از راهبردهای مؤثر در ارائه تصویری وضعیت تحلیل پیوند در محیط وب می باشد (شکل ۲). توزیع درجه ورودی^۳، و نیز درجه خروجی^۴، در شکل ۲ نشان داده شده است. هر دو توزیع به صورت سازگار، با بهره گیری از قاعده توان^۵ به انجام می رسند. هنگام آزمون توزیع درجه خروجی، دو منحنی متفاوت را ملاحظه می کنیم که هر دوی آنها نشان دهنده توزیع قاعده توان هستند و ما می توانیم ارزیابی و شرحی برای هر دو قسمت به صورت مجزا داشته باشیم (Baeza-Yates and Castillo 2005). این شیوه بیان تصویری در واقع با تحلیل همزمان فعالیت های فرایوند، همراه است. در این نمودار، معیارهای چندگانه در کنار هم نشان داده شده است. در سنجش و ارزیابی فرایند تحلیل پیوند، این نمودار از محاسبات آماری بهره می گیرد. در واقع محاسبه تحلیل پیوند در ارتباط با تأثیری که بر دیگر متون موجود در محیط وب برجای گذاشته است مورد ارزیابی قرار می گیرد، و

¹ navigational links

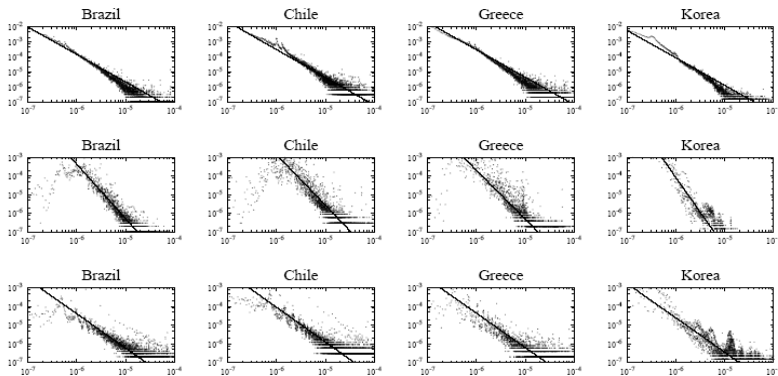
² recommendation links

³ indegree

⁴ outdegree

⁵ power-law

مقیاس‌های آماری با بیان ریاضی، وضعیت اثرپذیری پیوندها را معین می‌سازد. در شکل ۲ با مقایسه پیوندهای واقع شده در محدوده آثار علمی چند کشور، امتیازهای ارائه شده با ارقام نشان داده شده است. این نمودار در واقع با تغییرهای اعمال شده در تمامی گره‌های به‌نمایش درآمده در صفحه‌های مختلف وب مرتبط با وب‌سایت مادر خود، توسط بازنمون گره‌های واحد وب‌سایت ایجاد شده است. این گره‌ها می‌تواند برای تمامی وب‌سایت‌های حاوی اطلاعات علمی به کار رود. توزیع هر دو درجه و بررسی دقیق آن‌ها نشان می‌دهد که فعالیت علمی هر کشور تا چه حد مستند بوده است. همچنین در این قسمت باید به معیار مهمی که در این عرصه نقش بسزایی برای تبیین وضعیت مجموعه‌های علمی وب دارد اشاره کرد و آن، معیار رتبه-صفحه است.



شکل ۲. یک نمودار هیستوگرام از امتیازها در بهره‌گیری از راهبرد رتبه-صفحه (بالا)، محور

فعالیت (وسط)، و اعتبار (پایین) (Baeza-Yates and Castillo 2005)

همچنان که می‌دانیم، یکی از الگوریتم‌هایی که برای رتبه‌بندی پیوندها در محیط وب استفاده می‌شود، معیار رتبه‌صفحه می‌باشد. برای مجموعه‌های چندگانه، توزیع معیار رتبه-صفحه را مورد بهره‌برداری قرار می‌دهیم، و قاعده‌ای توان در توزیع امتیازهای فراهم آمده را با توان میانگین ۱:۸۶-۰:۶ ایجاد می‌کنیم. به صورت نظری توان رتبه-صفحه می‌بایستی شبیه به توان درجه ورودی باشد (که ارزش آن بر اساس توانی که ۱/۲ بوده

¹ Link-based ranking

محاسبه شده است) توزیع مقادیر رتبه- صفحه در شکل ۲ مشاهده می‌شود (Baeza-Yates and Castillo 2005). میانگین لبه‌های جهت‌دار در هر وب‌سایت بر اساس قانون توزیع زیف با در نظر گرفتن مقادیر پارامتری $0/3+1/8$ محاسبه شده است. در این شرایط میانگین تعداد وب‌سایت‌ها، درون همان کشوری که مرتبط با وب‌سایت معین در نظر گرفته شده، بررسی شده است. این شیوه سبب می‌شود مطالعه ویژگی‌های تحلیل پیوند، با دقت و صحت بالاتری به انجام رسد.

علاوه بر نمایش وضعیت رتبه- صفحه در نمودار وب، ما شاهد نمایش محور فعالیت در عرصه تولید علمی نیز هستیم. بر مبنای این بخش از نمودار، هسته‌های اصلی فعالیت علمی و اثربخشی آن‌ها به نمایش درآمده است. این بخش نیز با توزیع درجه ورودی و خروجی، به تبیین وضعیت تصویری از این مقوله پرداخته است. میانگین اصلی آن نیز با تعیین حدود محاسبه شده، مشخص است. ولی در قسمت سوم این نمودار اعتبار پیوندها با در نظر گرفتن امتیاز مربوطه آمده است. تعیین این امتیازها کمک می‌کند بتوانیم به شیوه مناسب‌تری به بررسی اعتبار پیوندها بپردازیم و با تکیه بر آن، به نحو بهتری به درک مراحل قبلی نائل شویم.

باید بیان کرد که سنجش اعتبار از حساسیت زیادی برخوردار است. با توجه به بررسی فرایند پیوند در محیط وب، «میزان اعتبار» از عواملی است که می‌تواند در ارائه نتایج کلی حاصل از بررسی پیوند، به پژوهشگران کمک کند و مراحل مطالعه تحلیل پیوند را کامل و معنادار سازد. در هر یک از موارد نامبرده، ما با میزان امتیازهای تعلق‌یافته روبرو هستیم. این امتیازها بر مبنای محاسبه خاص روش‌های آماری تحلیل پیوند صورت می‌گیرد که در نهایت در نمودار هیستوگرام ظاهر می‌شود و بررسی تحلیل پیوند را برای پژوهشگران آسان می‌سازد.

در اینجا ما نسخه امتیاز فرایند حاصل از جستجوی موضوعی را مورد سنجش قرار می‌دهیم. در ادامه فقط پیوندهای خارجی و امتیازهای نمودار کامل را در مقابل مجموعه کامل صفحات محاسبه می‌کنیم. به دنبال توزیع اعتبار، تبعیت ما از فرایند قاعده توان می‌باشد. در موقعیت امتیاز محور فعالیت، دشوار است که بتوان اثبات کرد که داده‌ها از قاعده توان استنباط می‌شوند، زیرا به نظر می‌رسد که بسامد آن به میزان زیادی از هم گسیخته است. ارزش توان میانگین محور فعالیت ۰:۳-۵:۵ مشاهده شده، و ۰:۱-۱:۴۸ در

مورد امتیاز اعتبار می‌باشد (Baeza-Yates and Castillo 2005). بر این اساس ما می‌توانیم گرافی را در نظر بگیریم که هر صفحه آن در حکم یک گره می‌باشد و فرایوند، میان دو صفحه یک لبه قرار گرفته است. این گراف، وب و فضای شبکه را در مقیاس وسیع شکل می‌دهد. یعنی گرافی که کار توزیع درجه گره‌هایی را که مورب است به انجام می‌رساند. این گراف همچنین خود-شبه‌ساز است، به این معنا که در نقاطی که با بخش کوچکی از گراف، مشترک است، می‌تواند ارتباط بیش تری با گراف اصلی برقرار سازد و در نتیجه امکان تطابق بالاتری در سنجش وضعیت پیوندهای موجود در فضای وب داشته باشد. این معیارهای سنجش که در شکل ۲ ارائه شده، یکی از ابزارهای سودمند برای اندازه‌گیری و ارزیابی معیارهای مؤثر بر فرایندهای تحلیل پیوند محسوب می‌شود. این نمونه‌ای سودمند در استفاده از نمودار مصور وب مربوط به ارائه وضعیت تحلیل پیوند در فضای وب است که به میزان زیادی در درک محاسبات استنادی پیوندها به محققان این حوزه موضوعی کمک می‌کند.

۹. نتیجه‌گیری

بررسی‌های کتاب‌سنجی در پوشش سنجش و محاسبات آماری وضعیت انتشارات در محیط وب، کاربردی مؤثر دارد. در تحلیل‌های کتاب‌سنجی از سه شاخص عمده استفاده می‌شود: نخست شمارش تعداد انتشارات، سپس عامل تأثیر و استناد، و در انتها تحلیل هم‌استنادی و هم‌واژه. این عوامل می‌تواند در بررسی و ارزشیابی متون اطلاعاتی در محیط وب به پژوهشگران و مؤسسه‌های مسئول در حوزه سنجش مطالب علمی کمک می‌کند تا شاخص‌های مؤثر ارزشیابی متون علمی را به‌نحو مناسبی به انجام رسانند. امروزه با توسعه و پیشرفت نرم‌افزارها و فناوری جدید برای تعیین رابطه میان عناصر اطلاعاتی موجود در مدارک محیط وب، می‌توان به ارائه شاخص‌های روابط میان اجزای اطلاعاتی، و تأثیر آنها بر تولید اطلاعات جدید پرداخت. از سوی دیگر این امکان وجود دارد که با کاربردهای متنوع فنون تحلیل استنادی به بررسی دقیق تأثیرپذیری اطلاعات موجود در محیط شبکه پردازیم.

در این میان یکی از عوامل مهم در بررسی اطلاعات موجود در فضای شبکه و اینترنت، فرایوندهای صفحات وب از عواملی هستند که متخصصان

موضوعی را در این عرصه یاری می‌کنند و می‌توان با طرح پرس و جوهای مشخص، به بررسی میزان ربط میان آن‌ها اقدام کرد. برای مطالعه در حوزه اطلاعات دیجیتال موجود در محیط وب و بررسی دقیق ارتباطات علمی، چندین ابزار مناسب وجود دارد. یکی از آن‌ها الگوی فرامتن مکانی است که سبب انتخاب بهینه موضوع می‌شود و کاربرد مفیدی در سنجش مربوط به حوزه تحلیل پیوند دارد.

ابزار دیگری که می‌تواند مراحل بررسی را تسریع کند، الگوریتم‌های تحلیل پیوند است که از جمله فنون کاربردی و سودمند آن، می‌توان به راهبرد رتبه-صفحه و فرایوند حاصل از جستجوی موضوعی در محاسبات بردار ویژه اشاره کرد، که برای انتخاب اطلاعات مستند سایت‌های معتبر به عنوان پایه‌ای در ساختار فرایوند، مورد استفاده قرار می‌گیرند.

نمایه‌سازی معنایی پنهان، دیگر ابزار تحلیل پیوند محسوب می‌شود که دارای توانایی الگوریتمی برای نمایه‌سازی و ساخت عبارت‌ها در صفحات و نیز مشخص ساختن موضوع آن صفحات است. این روش شامل تحلیل و بررسی واژه‌های به کار رفته در زبان طبیعی می‌شود. بر همین اساس، این راهبرد می‌تواند مترادف‌ها و واژگان مرتبط و وابسته را معین سازد تا وقتی که موضوع‌های کلی در یک صفحه وب مورد استفاده قرار می‌گیرند، بتواند به جای آن که واژگان را تعویض کند، آن‌ها را کامل سازد و در واقع تحلیل مناسبی از کلیدواژه‌ها ارائه دهد.

ابزار دیگری که دارای سودمندی بالایی است، مدل درون‌داد-برون‌داد می‌باشد که در تولید اطلاعات علمی، یکی از روش‌های مناسب برای شرح و توصیف فرایند پژوهش علمی است. بر اساس کارآمدی این مدل، تعیین مقیاس اندازه‌گیری در رابطه با استفاده از منابع علمی انجام می‌شود و سپس تولید عناصر اطلاعاتی بر مبنای سنجش ورودی-خروجی محاسبه می‌شود. باید اظهار داشت که این ابزار می‌تواند در حکم معیاری اثربخش و نسبتاً مطمئن در این حوزه مطرح شود. تمامی ابزارهای نامبرده می‌توانند به تقویت مطالعات علمی در حوزه تحلیل پیوند یاری رسانند و زمینه مناسبی برای بررسی‌های تحلیلی فراهم سازند. علاوه بر آن، چنانچه ناشران محیط وب بتوانند برون‌داد علمی خود را در قالب روشی مناسب ارائه دهند، شرایط مطلوبی برای سنجش‌های مربوط به حوزه تحلیل پیوند فراهم می‌آید. این بررسی‌ها معمولاً در دو محیط وب جهانی و وب کوچک

انجام می‌پذیرد. بر اساس فرضیه‌ای معتبر در این حوزه مطالعاتی، الگوریتم تحلیل پیوند در توصیف وب جهانی دارای یک نمودار یکپارچه استنادی است و هر فرایوند نمایش داده‌شده، یک استناد یا ارتباط توصیه‌شده به همراه دارد. ولی این فرضیه پیشنهادی در مورد محیط وب کوچک قابل سنجش نیست؛ زیرا وب کوچک بر اساس نوع اطلاعات، و همگنی میان آن‌ها، با خصوصیات متفاوتی ظاهر شده است.

همچنین برای ارائه تصویری ارتباطات میان فرایوندهای ارائه‌شده در محیط وب می‌توان از نمودارهای وب استفاده کرد که سبب درک بهینه از روابط ایجادشده در میان گره‌های وب، و پاره‌های اطلاعاتی می‌شود. نتایج این بررسی نشان می‌دهد که در ارتباط با کاربرد نمودارهای وب، نمایش دو توزیع درجه ورودی و خروجی می‌تواند معیارهای چندگانه‌ای را پشتیبانی کند و میزان درک بررسی از فرایند تحلیل پیوند را به‌شيوه مصور و بازنمون دیداری ارائه دهد. در واقع نمایش وضعیت رتبه‌صفحه در این نمودار، راهبردی برای سنجش شاخص‌های مهم فعالیت‌های علمی است. این مسئله با ارائه یک عنصر دیگر، یعنی محور فعالیت در عرصه تولید علمی کامل می‌شود. تمامی اجزای اطلاعاتی مربوط به فرایند تحلیل پیوند که در نمودارهای آن ظاهر می‌شوند، در یک جامعی از بررسی منابع علمی موجود در وب و رابطه استنادی میان اطلاعات علمی ارائه می‌دهند.

۱۰. فهرست منابع

- Archambault, Éric and Étienne Vignola Gagné. 2004. *Science-Metri: The Use of Bibliometrics in the Social Sciences and Humanities*. Final report. Prepared for the Social Sciences and Humanities Research Council of Canada (SSHRCC). www.science-metrix.com/pdf/Science-Metrix_Use_Bibliometrics_SSH.pdf (accessed Oct. 10, 2006).
- Baeza-Yates, Ricardo and Carlos Castillo. 2005. Link Analysis in National Web Domains. www.emse.fr/OSWIR05/2005-oswir-p15-baeza_yates.pdf (accessed Dec. 17., 2006).
- Bertin, M.J.P. Descl'es, B. Djioua and Y. Krushkov. 2006. *Automatic Annotation in Text for Bibliometrics Use*. Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference. Edited by Geoff Sutcliffe and Randy Goebel. Melbourne Beach, Florida. May 11–13, California: The AAAI Press, Menlo Park.
- Borgman, Christine L. and Jonathan Furner. 2002. forthcoming. Scholarly Communication and Bibliometrics. *Annual Review of Information Science and Technology* 36. www.infotoday.com/books/asist/arist36/sample.pdf (accessed Oct.10, 2006)
- Ding, Chris H.Q. 2003. Link analysis: Hubs and authorities on the world wide web. *LBNL Tech Report* www.crd.lbl.gov/~cding/papers/hits5.pdf (accessed Oct. 10, 2006).
- Glanzel, W. 2003. Bibilometrics as a research field: A course on theory and application of bibliometric indicators. *SIM Review* www.norslis.net/2004/Bib_Module_KUL.pdf (accessed Oct. 10, 2006).

- Jacobs, Daisy. 2001. A bibliometric study of the publication patterns of scientists in South Africa 1992-96, with particular reference to status and funding. *Information Research* 6(3). <http://InformationR.net/6-2/paper104.html> (accessed Oct.: 8, 2006).
- Kohlschütter, Christian, Paul-Alexandru Chirita, and Wolfgang Nejdl. 2006. Prototype Demonstration: Using Link Analysis to Identify Aspects in Faceted Web Search. www.l3s.de/~chirita/publications/kohlschuetter06using.pdf (accessed Dec. 17, 2006).
- Kolda, Tamara; and Brett Bader. 2006. *The TOPHITS Model for Higher-Order Web Link Analysis*. www.siam.org/meetings/sdm06/workproceed/Link%20Analysis/21Tamara_Kolda_SIAML_ACS.pdf (accessed: Dec. 17, 2006).
- Kolda, Tamara G., Brett W. Bader and Joseph P. Kenny. 2005?. Higher-Order Web Link Analysis Using Multilinear Algebra. csmr.ca.sandia.gov/~tgkolda/pubs/ICDM05.pdf (accessed Dec. 17, 2006).
- Larson, R.R. 1996. Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace. In *96: Global complexity: Information, chaos, and control: Proceedings of the 59th ASIS Annual Meeting Baltimore, MD*, Editor: S. Hardin, 71-78. USA: Baltimore, M. D.
- Lundberg, Jonas. 2006. *Bibliometrics: Publication Analysis as a Tool for Science Mapping and Research Assessment*. The Karolinska Institutet Bibliometrics Project Group. ki.se/content/1/c6/01/79/31/Introduction_to_bibliometrics_v1.11.pdf (accessed Oct. 10, 2006).
- Nisbet, Peter. 2006. Latent Semantic Indexing: What Is It? <http://www.aj2000.com/articles/latent-semantic-indexing.htm> (accessed: Dec. 19, 2006).
- Russell, Jane M. and Ronald Rousseau. 2002. Bibliometrics and Institutional Evaluation. www.vub.ac.be/BIBLIO/itp/lecturers/ronald_rousseau/ronald_rousseau_stim1_bibliometrics_russell.pdf (accessed: Dec. 11, 2006).
- Schiffman, Allan M. 2006. Hierarchy in Web Page Similarity Link Analysis. Commerce Net Labs Technical Report 06-02. <http://wiki.commerce.net/images/9/9c/CN-TR-06-02.pdf> (accessed Dec. 17, 2006).
- Sternberg, Sara Von Ungern. 1995. *Applications in teaching bibliometrics*. www.cindoc.csic.es/cybermetrics/pdf/441.pdf (accessed Oct. 10, 2006).
- Turnbull, Don. 1997?. *Bibliometrics and the World Wide Web*. <http://www.ischool.utexas.edu/~donturn/research/bibweb.html> (accessed Oct. 10, 2006).
- White, Howard D. 2001. Author-centered bibliometrics through CAMEOs: Characterizations automatically made and edited online. *Scientometrics* 51(3): 607-637.
- Wikipedia. 2008. Bibliometrics. <http://en.wikipedia.org/wiki/Bibliometrics> (Accessed Oct. 6, 2008).
- Xue, Gui-Rong, Hua-Jun Zeng, Zheng Chen, Wei-Ying Ma, Hong-Jiang Zhang and Chao-Jun Lu. 2003. *Implicit Link Analysis for Small Web Search*. Presented in SIGIR, Toronto, Canada. <http://research.microsoft.com/~zhengc/papers/p31261-xue.pdf> (accessed Dec. 17, 2006).

Web Applications of Bibliometrics and Link Analysis

Fariborz Doroudi*

PhD in Library and Information Science, National Library and Archives of Iran

Abstract: The present study aims to introduce and analyze bibliometric application within Web and also to expounds on the status of link analysis in order to point out its application with respect to the existing web-based information sources. Findings indicate that bibliometrics could have required application in the area of digital resources available through Net. Link analysis is a process by which one could make statistical analysis of correlation between hyperlinks and therefore understand the accuracy, veracity and efficacy of citations within a digital document. Link analysis, in effect, is counted as a part of information ranking algorithm within the web environment. The number, linkage and quality of given links to a website are of utmost importance for ranking/status in the Web. The tools applied in this topic include, page ranking strategy, link analysis algorithm, latent semantic indexing and the classical input-output model. The present study analyzes Big Web and Small Web link analysis and explains the means for utilizing web charts in order to better understand the link analysis process.

Keywords: bibliometrics; link analysis; webometrics; ranking; World Wide Web

* fardoroudi@gmail.com